

THE RANDOM TRANSPOSITION DYNAMICS ON RANDOM REGULAR GRAPHS AND THE GAUSSIAN FREE FIELD

SHIRSHENDU GANGULY AND SOUMIK PAL

ABSTRACT. A single permutation, seen as union of disjoint cycles, represents a regular graph of degree two. Consider d many independent random permutations and superimpose their graph structures. It is a common model of a random regular (multi-) graph of degree $2d$. We consider the following dynamics. The dimension (i.e. size) of each permutation grows by coupled Chinese Restaurant Processes, while in time each permutation evolves according to the random transposition chain. Asymptotically in the size of the graph one observes a remarkable evolution of short cycles and linear eigenvalue statistics in dimension and time. In dimension, it was shown by Johnson and Pal [JP14] that cycle counts are described by a Poisson field of Yule processes. Here, we give a Poisson random surface description in dimension and time of the limiting cycle counts for every d . As d grows to infinity, the fluctuation of the limiting cycle counts, across dimension, converges to the Gaussian Free Field. In time this field is preserved by a stationary Gaussian dynamics. The laws of these processes are similar to eigenvalue fluctuations of the minor process of a real symmetric Wigner matrix whose coordinates evolve as i.i.d. stationary stochastic processes.

1. INTRODUCTION

We begin with a heuristic description of our model and the main results. Precise formulations are given in the following subsection. Consider d independent random permutations on n labels $[n] := \{1, 2, \dots, n\}$. Every permutation has a corresponding permutation matrix whose entries are zero or one. We add all the d matrices and further add it to its own transpose. This produces a symmetric matrix such that the sum of entries in every row is exactly $2d$. The matrix can be thought of as the adjacency matrix of a regular (multi)graph that allows loops and multiple edges. The model of the random graph generated by this procedure is called the permutation model of a random regular graph of degree $2d$ on n vertices. We will denote this random graph by $G(n, 2d)$. For more on the recent uses and applications of the permutation model see [Fri91, Fri08].

Our objective in this paper is to consider stochastic processes of such random regular graphs indexed by two parameters: size (or ‘order’ or ‘dimension’) and time. The inspiration comes from a standard model of random matrix theory: a Gaussian Wigner matrix called the Gaussian orthogonal ensemble (GOE). In that model, at a given time and dimension n , one has an $n \times n$ symmetric matrix of independent

Date: December 30, 2014.

2000 Mathematics Subject Classification. 60B20, 60C05.

Key words and phrases. Random regular graphs, Chinese Restaurant process, Random transpositions, virtual permutations, Gaussian free field, minor process, Dyson Brownian motion.

Soumik’s research is partially supported by NSF grant DMS-1308340.

mean-zero Gaussian upper triangular entries (all entries have variance one, except the diagonal elements which have variance two). The matrix grows to dimension $(n+1) \times (n+1)$ by adding an independent $(n+1)$ th row of independent Gaussian entries and, hence, a column, by symmetry. This gives us a matrix-valued process indexed by n (called the minor process). If, now, the individual entries are replaced by independent Gaussian stochastic processes (say, stationary Ornstein-Uhlenbeck diffusions), then we have a field of random matrices indexed by size and time that is known to display remarkable properties.

The corresponding field of random regular graphs is non-trivial since the entries of the adjacency matrix are not independent. However, each graph is constructed using random permutations. One can borrow well-known dynamics on every random permutation: the Chinese Restaurant Process (CRP) to grow its size and the random transposition Markov chain to evolve it in time. Dynamics on individual permutations must be coupled to ensure that the same set of labeled vertices are preserved in dimension and time.

To do this, imagine observing an infinite sequence $\{\pi_1, \pi_2, \dots\}$ of independent random permutations of \mathbb{N} . By a ‘random permutation’ of \mathbb{N} we mean the following. The i th element π_i is a sequence of permutations visualized as a tower. The n th level of this tower is a random permutation of $[n]$ that grows in n according to the CRP. This, in turn, produces an array of permutation matrices doubly indexed by (i, n) . Consider the sequence of partial sums of these permutation matrices along i for every n . Symmetrize every matrix by adding it to its transpose. Consider the d th partial sum and consider the process of matrices growing in n . The corresponding sequence of graphs produces a coupling of $G(n, 2d)$ for all values of n .

We now describe the evolution in time. Attach each element of \mathbb{N} with a sequence of i.i.d. exponential clocks. When any clock rings the corresponding element chooses a ‘uniform’ element from \mathbb{N} and every permutation tower gets multiplied on the left by the transposition of the pair. Clearly this statement as it is does not make sense since there are countably infinitely many elements. However, as we show later, there is a way to make this precise. The evolving family of towers of permutations now produces a family $G(n, 2d, s)$ of regular graphs indexed by the triplet (order, degree, time). Consider the following graph statistics (i) counts of cycles of a fixed size, (ii) polynomial linear eigenvalue statistics. We study their functional limits for large order and suitably scaled time both for fixed d and as d tends to infinity.

For very large order, the cycle counts form an approximate polynomial basis for the linear eigenvalue statistics which makes (i) and (ii) asymptotically equivalent. We provide a precise process description of the joint evolution of cycles of various sizes across dimension (i.e., order) and time. Informally, suppose $(N_k(t, s), k \in \mathbb{N})$ denote the number of k cycles in the graph for *very large* order t and *very small* time s , then its joint law is approximately that of countably many random Poisson surfaces (one per k) given by Yule processes in dimension and approximate birth-and-death chains in time. The dimension part of it is described in [JP14]. As d goes to infinity, the fluctuation of the above infinite dimensional surface converges to a product of countably many two dimensional Gaussian random surfaces (one for each k) each of which has stationary one dimensional marginals. The covariances of these random surfaces coincide with that of Chebyshev polynomial eigenvalue statistics of Wigner matrices, growing in dimension as the minor process, while each entry (up to symmetry) moves in time as independent processes. See

[Bor10, Bor14]. It follows that the fluctuation of the height function of eigenvalue distribution (for every fixed time) is distributed approximately as the Gaussian Free Field (GFF) on the upper half plane and zero boundary condition, whose law is kept preserved in time by the random transposition Markov chain.

1.1. Formal description of the model. A part of the description of our model already appears in [JP14] and [Joh14], where the reader can find more references on the subject. Consider a permutation $\pi^{(n)}$ on the n labels $[n] := \{1, 2, \dots, n\}$. Consider the permutation matrix corresponding to $\pi^{(n)}$, add it to its transpose. The resulting matrix can be thought of as the adjacency matrix of a vertex labeled 2-regular graph that allows multiple edges and loops.

We will now grow this graph in dimension and transform it in time. In dimension this will be done by the Chinese restaurant Process (CRP). By a tower of random permutations we mean a sequence of random permutations $(\pi^{(n)}, n \in \mathbb{N})$ such that

- (i) $\pi^{(n)}$ is a uniformly distributed random permutation of $[n]$ for each n , and
- (ii) for each n , if $\pi^{(n)}$ is written as a product of cycles then $\pi^{(n-1)}$ is derived from $\pi^{(n)}$ by deletion of the element n from its cycle.

The CRP is a Markov chain that reverses the above procedure by building $\pi^{(n)}$ from $\pi^{(n-1)}$. See [Pit06, Section 3.1].

Now suppose we construct a countable collection $\{\Pi_d, d \in \mathbb{N}\}$ of towers of random permutations. We will denote the permutations in Π_d by $\{\pi_d^{(n)}, n \in \mathbb{N}\}$. Then it is possible to model every possible $G(n, 2d)$ by adding the permutation matrices (and their transposes) corresponding to $\{\pi_j^{(n)}, 1 \leq j \leq d\}$. In what follows we will keep d fixed and consider n as a growing parameter. Thus, G_n will represent $G(n, 2d)$ for some fixed d . Here and later, G_0 will represent the empty graph. We construct a continuous-time version of a graph-valued Markov chain by inserting new vertices into G_n with rate $n + 1$. Formally, define independent times $T_i \sim \text{Exp}(i)$, exponential distribution with rate i , and let

$$(1) \quad M_t = \max \left\{ m : \sum_{i=1}^m T_i \leq t \right\},$$

and consider the continuous-time Markov chain G_{M_t} for $t \in [0, \infty)$. We are now going to abuse our notation and define a two-parameter family of graphs $G(t, s)$, for nonnegative parameters t and s . The second index s in $G(t, s)$ refers to time, while the first index t refers to dimension. This should not produce confusion with the notation $G(n, 2d)$, where the parameters are integers and d is fixed. We start by defining $G(t, 0) = G_{M_t}$ for all $t \geq 0$.

We now describe the movement in time. Fix some positive T and suppose $M_T = n$. Let τ_{ij} be the transposition (i, j) . Consider the finite set of transpositions of elements in $[n]$. Consider the Markov chain that independently chooses a uniform random transposition and multiplies to each permutation $\pi_i^{(n)}$ on the left. This is the well-known random transposition Markov chain which keeps the joint law of d independent permutations $\{\pi_i^{(n)}, i \in [d]\}$ invariant.

As before, we will actually modify the above to a continuous time chain. Suppose we attach a sequence of i.i.d. exponential one clocks with every label in $[n]$. As the first clock rings, the corresponding element (say I_1) chooses a uniformly random element in $[n]$ except itself (say J_1) and we multiply every permutation

$\{\pi_j^{(n)}, j \in [d]\}$ on the left by $\tau_{I_1 J_1}$. With every successive ring one takes a successive product $\tau_{I_k J_k} \cdot \dots \cdot \tau_{I_1 J_1} \cdot \pi_j^{(n)}$.

Thus every possible transposition occurs with rate $2/(n-1)$ and we successively multiply them on the left. After time s , let σ_s denote the (left) product of successive transpositions so far. Then, each permutation is now modified to $\sigma_s \cdot \pi_j^{(n)}$. The graph induced by these permutations will be denoted by $G(T, s)$.

Given initial permutations $\{\pi_i^{(n)}, i \in [d]\}$ and σ_s for some $s \geq 0$, we define $G(t, s)$ for $t \in [0, T]$ by successively removing elements in the order

$$(2) \quad \sigma_s(n), \sigma_s(n-1), \dots, \sigma_s(1).$$

More precisely, recall the sequence of exponential times T_1, T_2, \dots, T_n that defined $G(T, 0)$ in (1) with $M_T = n$. Consider the CRP backwards at time 0 as it removes vertices from each permutation. When it removes vertex i at time 0, simultaneously remove vertex $\sigma_s(i)$ from at time s . This is the CRP running backwards on the relabeled vertices (2). Hence, the law of the unlabeled graph-valued process in dimension remains unchanged along time. Additionally, by the above coupling every $G(t, s)$, $s \geq 0$, has the same number of vertices for all $0 \leq t \leq T$, although their labels might be different.

Remark. Recall the comment made in the introduction about the difficulty in defining transposition Markov chain on a permutation on the entire \mathbb{N} . We get around this problem by defining the transposition chain for a permutation of a large dimension T and then project back to smaller dimensions by running the CRP backwards. We will later take $T \rightarrow \infty$ to construct a substitute for the entire \mathbb{N} .

For every $T > 0$ the above construction produces a doubly-indexed family of graphs $\{G(t, s), 0 \leq t \leq T, s \geq 0\}$, where t represents dimension and s represents time. We intend to study the asymptotic behavior of short cycles and linear eigenvalue statistics of this process as T grows to infinity. For a fixed time, the asymptotic law of the process in dimension has been already studied in [JP14], to be described in the next section. The main focus of this work is to study the joint evolution of the graph statistics in dimension and time and to draw parallel with results in [Bor10].

1.2. Notation and Definitions. We are interested in the dynamics of cycles of a given size. However, not all cycles of a fixed size behave identically. To obtain a nice Markovian description, we need to classify cycles by the permutations that produce its edges. The following concepts are recalled from [JP14].

Imagine the graph G_n as a directed, edge-labeled graph in a natural way. For convenience, drop superscripts and let $\pi_l = \pi_l^{(n)}$. If $\pi_l(i) = j$, then we imagine this edge to be directed from i to j and to be labeled by π_l .

Consider a walk on G_n (i.e., a sequence of neighboring vertices) and write down the label of each edge as it is traversed, putting π_i or π_i^{-1} according to the direction we walk over the edge. Any such sequence of π_i or π_i^{-1} will be called a word.

We call a walk *closed* if it starts and ends at the same vertex, and we call a closed walk a cycle if it never visits a vertex twice until the very last one, and it never traverses an edge more than once in either direction. Thus a word $w = w_1 \dots w_k$ that corresponds to traversing a cycle is cyclically reduced, i.e., $w_i \neq w_{i+1}^{-1}$ for all i , considering i modulo k .

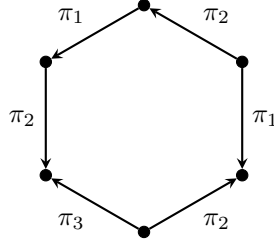


FIGURE 1. A cycle whose word is the equivalence class of $\pi_2\pi_1^{-1}\pi_2\pi_1\pi_2\pi_3^{-1}$ in \mathcal{W}_6/D_{12} . Here $h(w) = 1$, $b(w) = 2$, $c(w) = 0$.

Let \mathcal{W}_k denote the set of cyclically reduced words of length k . We identify elements of \mathcal{W}_k that differ only by rotation and inversion and denote the resulting set by \mathcal{W}_k/D_{2k} , where D_{2k} is the dihedral group acting on the set \mathcal{W}_k in the natural way. Let $\mathcal{W}' = \bigcup_{k=1}^{\infty} \mathcal{W}_k/D_{2k}$, and let $\mathcal{W}'_K = \bigcup_{k=1}^K \mathcal{W}_k/D_{2k}$. For each k -cycle in G_n we associate an element in \mathcal{W}_k/D_{2k} formed by starting the walk at any point in the cycle and walk in either of two directions. Two cycles are considered equivalent if they both map to the same equivalent class of words.

Definition 1 (Properties of words). For any k -cycle in G_n , the element of \mathcal{W}_k/D_{2k} given by walking around the cycle is called the *word* of the cycle (see Figure 1). For any word w , let $|w|$ denote the length of w . Let $h(w)$ be the largest number m such that $w = u^m$ for some word u . If $h(w) = 1$, we call w *primitive*. For any $w \in \mathcal{W}_k$, the orbit of w under the action of D_{2k} contains $2k/h(w)$ elements, a fact which we will frequently use. The sign of a letter in a word is $+1$ or -1 depending on whether the letter is π_i or π_i^{-1} . Let $b(w)$ denote the number of letters in w whose sign is the same as the letter appearing right before it. In other words, $|w| - b(w)$ is the number of successive sign changes in w . Let $c(w)$ denote the number of pairs of double letters in w , i.e., the number of integers i modulo $|w|$ such that $w_i = w_{i+1}$.

For example, if $w = \pi_1\pi_1\pi_2^{-1}\pi_2^{-1}\pi_1$, then $b(w) = 2$ and $c(w) = 3$. See Figure 1 for another example. We will consider $|\cdot|$, $h(\cdot)$, $b(\cdot)$ and $c(\cdot)$ as functions on \mathcal{W}_k/D_{2k} , since they are invariant under cyclic rotation and inversion.

To more easily refer to words in \mathcal{W}_k/D_{2k} , choose some representative $w_1 \cdots w_k \in \mathcal{W}_k$ for every $w \in \mathcal{W}_k/D_{2k}$. Based on this, we will often think of elements of \mathcal{W}_k/D_{2k} as words instead of equivalence classes, and we will make statements about the i th letter of a word in \mathcal{W}_k/D_{2k} . For $w = w_1 \cdots w_k \in \mathcal{W}_k/D_{2k}$, let $w^{(i)}$ refer to the word in $\mathcal{W}_{k+1}/D_{2k+2}$ given by $w_1 \cdots w_i w_i w_{i+1} \cdots w_k$. We refer to this operation as *doubling* the i th letter of w . A related operation is to *halve* a pair of double letters, for example producing $\pi_1\pi_2\pi_3\pi_4$ from $\pi_1\pi_2\pi_3\pi_4\pi_1$. Since we apply these operations to words identified with their rotations, we do not need to be specific about which letter of the pair is deleted.

Finally, we adopt the following convention regarding words. As our permutations get modified by multiplication with transpositions, we do not change the notation of the letters. That is, at time zero, the letters are $\{\pi_i, \pi_i^{-1}, 1 \leq i \leq d\}$ and they remain so for all time s regardless of the fact that each π_i is now modified to $\sigma_s \cdot \pi_i$. Therefore, the edge labels due to a word, say $\pi_1\pi_2^{-1}\pi_3\pi_1$, at any time s should be understood as being given by the resulting permutation at time s .

Consider a word $w \in \mathcal{W}_k/D_{2k}$, and let $C_w(t, s)$ denote the number of cycles with word w that exists in $G(t, s)$. We can now formulate a regime where the limit of the processes $C_w(\cdot, \cdot)$ can be described. Fix some $T > 0$ and consider the doubly indexed process $(C_w((T+t)_+, s), w \in \mathcal{W}', t \leq 0, s \geq 0)$ where $x_+ := \max(x, 0)$.

We will take limit as T goes to infinity to get a process

$$(3) \quad (N_w(t, s), w \in \mathcal{W}', t \leq 0, s \geq 0)$$

which is described below. We imagine that in dimension this process is indexed by the negative half-line with a *front* at zero. In dimension, this process will be stationary. Recall the following Markov chain from [JP14, Lemma 13].

Definition 2 (The halving chain). Consider a time homogenous Markov chain with RCLL paths on the state space $\mathcal{W}' \cup \{\Delta\}$, where Δ is the cemetery. Let $u \in \mathcal{W}_{k-1}/D_{2k-2}$ and $w \in \mathcal{W}_k/D_{2k}$ be two words such that u can be obtained from w by halving j different pairs of letters. The transition kernel of the chain is described below.

- (i) The chain jumps from w to u at rate j .
- (ii) The chain jumps from w to Δ (i.e., gets killed) at rate $(|w| - c(w))$.

The following definition encapsulates the following simple idea. At dimension 0, the number of cycles with word w is a birth-and-death chain running in stationarity. Once born, each cycle, looked backward in dimension performs the halving chain. If the same cycle exists at two different time points, then we observe an identical backward path. Different cycles behave independently.

Definition 3 (The limiting process). Consider a Poisson point process (PPP) χ on $(-\infty, \infty) \times [0, \infty) \times \mathcal{W}'$ with an intensity measure that can be described in the following way. For any word $w \in \mathcal{W}'$, atoms appear on $(-\infty, \infty) \times [0, \infty) \times \{w\}$ with a rate given by the product of Lebesgue measure on $(-\infty, \infty)$, the exponential probability measure of rate $2b(w)$ on $[0, \infty)$, and the delta mass $2b(w)/h(w)\delta_w$.

Consider an extension of this probability space to support independent halving chains starting from every atom of χ . For an atom labeled (z, v, w) , we will call this chain $X_{z,v,w}(u)$, where $u \geq 0$ is the common ‘time’ parameter for these chains. For $t \leq 0, s \geq 0$ and $\bar{w} \in \mathcal{W}'$, define the collection of random variables

$$(4) \quad N_{\bar{w}}(t, s) \triangleq \sum_{(z,v,w) \in \Lambda^*(t,s)} 1\{X_{z,v,w}(-t) = \bar{w}\}.$$

Here $\Lambda^*(t, s)$ is the collection of all atoms (z, v, w) in χ such that $z \leq s \leq z + v$. This is what we will refer to as the limiting process or the limiting field.

The birth of a cycle is captured by the first coordinate of any atom of χ , the second notes its lifetime, while the third is the word of the cycle.

We are now ready to formally state the main results in this paper.

2. MAIN RESULTS

Our first result describes marginal cycle counts. Here and below, the topology of weak convergence is a natural generalization of the Skorokhod topology in higher dimensions as described in [Neu71]. We will provide more details later.

Theorem 1. Fix any $T_0, S_0 > 0$. As T tends to infinity, the cycle counting field,

$$(C_w(T + t, s), \quad (t, s) \in [-T_0, 0] \times [0, S_0], \quad w \in \mathcal{W}'),$$

converges weakly in to the field $(N_w(\cdot, \cdot), \quad w \in \mathcal{W}')$ defined in Definition 3.

Fix any $k \in \mathbb{N}$. Define a family of random variables

$$(5) \quad N_k(t, s) := \sum_{w \in \mathcal{W}_k / D_{2k}} N_w(t, s), \quad t \leq 0, s \geq 0.$$

Thus, $N_k(t, s)$ counts the number of k cycles at dimension t and time s .

Let

$$(6) \quad a(d, k) = \begin{cases} (2d - 1)^k - 1 + 2d, & \text{when } k \text{ is even,} \\ (2d - 1)^k + 1, & \text{when } k \text{ is odd.} \end{cases}$$

It is shown in [DJPP12, Lemma 41] that $a(d, j)/2j$ is the size of \mathcal{W}_j / D_{2j} .

For every fixed (t, s) , the variable $N_k(t, s)$ has the limiting law of the number of k -cycles in random regular graph of degree $2d$. It follows from [DJPP12] that every $N_k(t, s)$ is Poisson with mean $a(d, k)/2k$. Let

$$(7) \quad X_k(t, s) = (2d - 1)^{-k/2} (2k N_k(t, s) - a(d, k)).$$

Then, X_k 's are centered random variables with variance one.

It has been shown in [JP14] that, for every fixed s , the vector-valued process $(X_k(\cdot, s), \quad k \in \mathbb{N})$ converges in law to a family of independent stationary Ornstein-Uhlenbeck (OU) processes. Here we show surface convergence. Recall that a Yule process ξ is a pure-birth process on \mathbb{N} with generator

$$Lf(k) = k(f(k+1) - f(k)), \quad k \in \mathbb{N}.$$

Theorem 2. As $d \rightarrow \infty$, the field $(X_k(t, s), \quad t \leq 0, s \geq 0, k \in \mathbb{N})$ converges weakly to a family of continuous Gaussian surfaces $(U_k(t, s), \quad t \leq 0, s \geq 0, k \in \mathbb{N})$ over any compact rectangle.

(i) If $j \neq k$, then U_j and U_k are independent.

(ii) Suppose $j = k$. Consider, two elements (t_1, s_1) and (t_2, s_2) in $(-\infty, 0] \times [0, \infty)$. If $\theta := -\max(t_1, t_2)$ and $s := |s_1 - s_2|$ then

$$(8) \quad \text{Cov}(U_j(t_1, s_1), U_j(t_2, s_2)) = 2j e^{-j|t_1 - t_2|} \left[\mathbb{E} \left(e^{-2s\xi(\theta)} \right) \right]^j \mathbb{E} \left(e^{2s\tau} \right),$$

where $\xi(\theta)$ is the state of a Yule process at time θ , starting from $\xi(0) = 1$, and τ is the random variable that counts the number of sign changes along a j -cycle if we attach random i.i.d. ± 1 at every vertex.

Our next results makes precise the idea of running time infinitesimally slowly.

Theorem 3. For every finite rectangle in $\mathbb{R} \times [0, \infty)$, consider T_0 large enough such the following process is well-defined:

$$\left(U_j \left(-T_0 + u, \frac{1}{2} v e^{-T_0} \right), \quad j \in \mathbb{N} \right)$$

where (u, v) lies in the rectangle. As T_0 tends to infinity, the weak limit of this field is another family of Gaussian surfaces $(G_j(u, v), \quad u \in \mathbb{R}, v \geq 0)$, independent for each j , with the following non-trivial covariance structure:

$$(9) \quad \text{Cov}(G_j(u_1, v_1), G_j(u_2, v_2)) = 2j \left(\frac{e^{-|u_1 - u_2|}}{1 + |v_1 - v_2| e^{-\max(u_1, u_2)}} \right)^j.$$

Note that, for fixed time parameter $v_1 = v_2$ the process is a stationary OU process. For a fixed dimension $u = u_1 = u_2$, the process is a stationary Gaussian process. In conclusion, this Gaussian field can be thought of as the asymptotic fluctuation of cycle counts in the heuristic set-up described in the very beginning.

We now focus on eigenvalues of $G(t)$. For any d -regular graph on n vertices G and function $f: \mathbb{R} \rightarrow \mathbb{R}$, define the random variable

$$\text{tr} f(G) := \sum_{i=1}^n \hat{f}(\lambda_i)$$

where $\lambda_1 \geq \dots \geq \lambda_n$ are the eigenvalues of adjacency matrix of G divided by $2(2d-1)^{1/2}$ and \hat{f} is f with the constant term adjusted. The details, similar to [JP14], will be described later. By a polynomial basis we refer to a sequence of polynomials $\{f_0 \equiv 1, f_1, f_2, \dots\}$ such that f_k is a polynomial of degree k of a single argument over reals.

Theorem 4. *There exists a polynomial basis $\{f_i, i \in \mathbb{N}\}$ (depending on d) such that for any $K \in \mathbb{N}, T_0 > 0, S_0 > 0$, the process*

$$(\text{tr} f_k(G(T+t, s)), k \in [K], t \in [-T_0, 0], s \in [0, S_0])$$

converges in law, as T tends to infinity, to the limiting field $(N_k(t, s), k, t, s)$ of Proposition 1. Hence, for any polynomial f , the process $(\text{tr} f(G(T+t, s)))$ converges to a linear combination of $(N_k(t, s), k, t, s)$.

For our final result we will take d to infinity. We will make the following notational convention: for any polynomial f , we will denote the corresponding linear combination from Theorem 4 by $\text{tr} f(G(\infty + t, s))$.

Theorem 5. *Let $\{T_k, k \in \mathbb{N}\}$ denote the Chebyshev orthogonal polynomials of the first kind on $[-1, 1]$. As d tends to infinity, the collection of processes*

$$(10) \quad (\text{tr} T_k(G(\infty + t, s)) - \mathbb{E}[\text{tr} T_k(G(\infty + t, s))], t \leq 0, s \geq 0, k \in \mathbb{N})$$

converges weakly to the Gaussian field $U/2$, as in Theorem 2. In particular, under the set-up of Theorem 3, (half of) the same weak limit holds.

The independent Gaussian evolution of Chebyshev polynomials in (9) is similar to Borodin's result on the dynamics of minor processes of stochastically evolving Wigner random matrices. See [Bor10, Proposition 3]. This explains how the random transposition chain preserves the GFF fluctuations of eigenvalues in a manner similar to Dyson Brownian motion. A more detailed comparison is given below.

2.1. Relevant existing literature and comparison with Wigner. In this subsection we discuss analogies between this work and [Bor10]. Some of the notations below is borrowed from that source. We keep our description informal. The reader is also encouraged to look at the introduction to [JP14]. Consider two families of independent identically distributed real-valued stochastic processes $\{Z_{ij}(t), j > i \geq 1, t \in \mathbb{R}\}$ and $\{Y_i(t), i \geq 1, t \in \mathbb{R}\}$ which have zero mean and suitable higher-moment conditions. Assume that there is a kernel $c(s, t)$ such that $c(s, t) \geq 0, c(t, t) = 1$ such that

$$\begin{aligned} \mathbb{E}(Z_{12}(s)Z_{12}(t)) &\equiv c(s, t) \equiv \frac{1}{2}\mathbb{E}(Y_1(s)Y_1(t)), \\ \mathbb{E}(Z_{12}^2(s)Z_{12}^2(t)) &\equiv 2c^2(s, t) + 1. \end{aligned}$$

Define a process of infinite Wigner matrices $X(t)$ by

$$X_{ij}(t) = \begin{cases} Z_{ij}(t), & i < j, \\ Y_i(t), & i = j, \\ X_{ji}(t), & i > j. \end{cases}$$

When the entry processes are standard Ornstein-Uhlenbeck processes on \mathbb{R} , the spectrum evolves as a process known as Dyson Brownian motion. In this setting at a single time point the matrix arising forms the Gaussian Orthogonal Ensemble (GOE) ($\beta = 1$).

For any $n \in \mathbb{N}$, let $X(n, t)$ denote the $n \times n$ principal submatrix of $X(t)$. Fix a parameter L that will be sent to infinity. Let z be a complex number in the upper half plane \mathbb{H} . Let $y = |z|^2$ and $x = 2\Re(z)$. The height function H_L of the eigenvalue distribution of $X(t)$ is a function indexed by $\mathbb{H} \times \mathbb{R}$ given by

$$H_L(z, t) = \sqrt{\frac{\pi}{2}} \# \left\{ \text{eigenvalues of } X(\lfloor Ly \rfloor, t) \geq \sqrt{L}x \right\}.$$

Then, Borodin shows that, as L tends to infinity, $\{H_L(z, t), z \in \mathbb{H}, t \in \mathbb{R}\}$, seen as a stochastic process of random distribution on \mathbb{H} , converges in law to a generalized Gaussian process on $\mathbb{H} \times \mathbb{R}$ whose every t marginal is the GFF on \mathbb{H} with zero boundary condition.

The law of this limiting Gaussian process can be characterized by Chebyshev polynomials ([Bor10, Proposition 3]). In short, consider Chebyshev polynomials $(T_j, j \in \mathbb{N})$. Consider the corresponding fluctuations of linear eigenvalue statistics:

$$\text{tr}(T_j(X(\lfloor Ly \rfloor, t))) - \mathbb{E}[\text{tr}(T_j(X(\lfloor Ly \rfloor, t)))].$$

Borodin shows, as $L \rightarrow \infty$, that the collection of limiting centered Gaussian surfaces, indexed by \mathbb{N} , are independent of one another. For a given j , two points on the surface (y_1, t_1) and (y_2, t_2) have a non-trivial covariance given by

$$(11) \quad \frac{j}{2} \left(\sqrt{\frac{y_1}{y_2}} c(t_1, t_2) \right)^j, \quad y_1 \leq y_2.$$

Consider now Theorem 4 and the final Gaussian field in Theorem 3 (divided by 2). Not only the Gaussian surfaces for different Chebyshev polynomials are independent, but for a given polynomial the covariance structure is almost identical to (11). There are two differences though. One, there is a re-parametrization of $y_i = e^{2u_i}$ which gives us

$$\sqrt{\frac{y_1}{y_2}} = e^{-|u_1 - u_2|}, \quad y_1 \leq y_2.$$

More importantly, if one takes the only possible choice of $c(v_1, v_2) = (1 + |v_1 - v_2|)^{-1}$, there is an additional term that does not match with (11).

However, if $u_1 = u_2 = u$ and we redefine v to ve^{-u} , it does give us the correct expression. The heuristic explanation is that while in the case of Borodin every minor process is moving in time at the same speed, for the random transpositions the speed depends on the size of the graph being considered.

Our analysis is also seemingly related to a series of work by [KOV04] and others, where a sequence of permutations satisfying condition (ii) is called a *virtual permutation*, and the distribution on virtual permutations satisfying condition (i) is

considered as a substitute for Haar measure on $S(\infty)$, the infinite symmetric group. Although many of the same ingredients appear in both these works, our work considers several permutations while the other considers a single infinite permutations.

Random matrix theory for sparse random regular graphs is a recent area of research which is not covered under the rubric of universality of traditional random matrix ensembles. See the discussion in the article [DP12]. However, empirical spectra distributions have been shown to approximate Wigner's semicircle law in different limiting regimes in [DP12] and [TVW13]. Both the above articles consider a sequence $\{d_n\}$ of degree that goes to infinity with the size n of the graph, although several results in [DP12] extend easily to the case of a fixed degree d . The study of linear eigenvalue statistics for both fixed d and growing $\{d_n\}$ is done in [DJPP12] where many similar combinatorial objects were exploited. The closest relative of the current article is [JP14] which covers the dimension dynamics for fixed d , much as the current article. The thesis [Joh14, Chapter 4] extends the ideas in [JP14] to growing $\{d_n\}$ where a complete proof of the convergence of fluctuation of height function to the GFF has been done.

Let us also mention that when $d = 1$, the process in dimension is a continuous-time version of the CRP itself while, in time, it is the well-known random transposition Markov chain. The latter has been studied in several contexts. See, for example, the references in the book [Dia88]. In particular, a long chain of literature is devoted to mixing properties of the chain. Modern bounds and more references can be found in the article [BSZ11]. A slightly related study is the effect this chain has on large cycles of the permutation (the split-merge transformation). See the article [DMWZZ04]. In our context the case $d = 1$ is unusual compared to $d > 1$. For example, $G(t, s)$ is likely to be disconnected when $d = 1$ and connected when d is larger. However, our results for finite d continue to hold.

3. PROPERTIES OF THE LIMITING FIELD

We start with the process described in Definition 3. The PPP χ and the countably many halving processes can clearly be constructed on a suitable probability space. What is not obvious is why the field $N(\cdot, \cdot)$ as defined in (4) is finite almost surely. In this subsection we prove this and other properties of the limiting field.

The following definition will be used throughout the rest of the article.

Definition 4. A few colloquial conventions regarding an atom (z, v, w) of χ or any other point process on the same space. We will refer to w as the word of the atom. We say that the atom is *born* during time interval J if $z \in J$. We say that the atom *exists* at time s if $z \leq s \leq z + v$. The middle coordinate v will be referred to as the lifetime of the atom. We will frequently use the memoryless property of the lifetime distribution without mention. Also, whenever we write 'time', it refers to the time of the halving chains, which is really the dimension running backwards for the limiting field.

Lemma 6. *For any $w \in \mathcal{W}'$, the process $(N_w(0, s), s \geq 0)$ is a continuous time birth-and-death chain on the state space $\{0, 1, 2, \dots\}$ and generator*

$$(12) \quad \frac{2b(w)}{h(w)} (f(x+1) - f(x)) + 2xb(w) (f(x-1) - f(x)) 1_{\{x>0\}}.$$

The chain is running in stationarity and is time-reversible. The collection of processes $(N_w(0, \cdot), w \in \mathcal{W}')$ are independent of one another. In particular, the distribution of $(N_w(0, s), w \in \mathcal{W}')$, for any fixed s , is the product measure of independent Poisson($1/h(w)$), $w \in \mathcal{W}'$.

Proof of Lemma 6. From the PPP structure, it is immediate that the processes $N_w(0, \cdot)$ are independent for various w . For a fixed w , atoms arrive at a rate $2b(w)/h(w)$ and survive an i.i.d. exponentially distributed amount of time. Clearly, $N_w(0, \cdot)$ is a continuous time Markov chain with generator (12). This is obviously a birth-and-death chain. Elementary arguments show that the unique stationary law is Poisson($1/h(w)$) under which it is reversible. \square

The above is a particular case covered in Proposition 10 stated later. However we defer stating it since one needs a few more definitions to do that.

Now recall (4) and the halving chain in Definition 2. The next lemma roughly states that given any finite (dimension, time) rectangle and words of length at most K , with high probability, there exists an L large enough such that all such words must have shrunk from words of length at most L at dimension zero.

Lemma 7. *Fix $K, L \in \mathbb{N}, L > K$, let χ_L denote the restriction of χ to $[0, \infty) \times [0, \infty) \times \mathcal{W}'_L$. Define the field*

$$N_{\bar{w}}^{(L)}(t, s) \triangleq \sum_{(z, v, w) \in \Lambda^*(t, s) \cap \chi_L} 1 \{X_{z, v, w}(-t) = \bar{w}\}.$$

Then, given any rectangle $R := [-T_0, 0] \times [0, S_0] \times \mathcal{W}'_K$, for $T_0, S_0 > 0$ and $K \in \mathbb{N}$ and any $\epsilon > 0$, there exists an $L \gg K$ such that

$$(13) \quad \mathbb{P} \left(N_w(t, s) = N_w^{(L)}(t, s), (t, s, w) \in R \right) \geq 1 - \epsilon.$$

In particular, the random variable $\sup_{(t, s, w) \in R} N_w(t, s)$ is almost surely finite.

Step 2 The proof of [JP14, Theorem 16] proves the one dimensional version of the above statement. For the benefit of the reader we use similar notations since the arguments are quite similar with necessary generalizations. The basis of the argument is careful counting followed by the union bound.

Proof of Lemma 7. Fix $L > K$. Consider a word w of size $l \geq L$ in an atom of χ born during $[0, S_0]$. Suppose a halving chain starting at w reduces to a word $\bar{w} \in \mathcal{W}'_K$ by ‘time’ T_0 . Then, it has to halve at least $l - K$ times during $[0, T_0]$. To bound this probability, we recall the transition kernel of the halving chain given in Definition 2. An easier description of the halving probability is to attach independent exponential clocks of rate 1 at every letter i such that $w_i = w_{i+1}$, modulo the length. These are the positions that can be halved. Whenever a clock rings, we erase that letter and the state of the chain has jumped to a new word. Of course, the chain can be killed at any time, but since we are only interested in upper bounds on cycle counts, we can ignore this event.

Let $E(l)$ be the event that for some word of length l existing in χ during $[0, S_0]$ jumps at least $l - K$ times by ‘time’ T_0 . For $l \geq L$, word $w \in \mathcal{W}_l/D_{2l}$, and $I \subseteq [l]$, let $F(w, I)$ denote the event that there exists an atom during $[0, S_0]$ of word w such that the halving chain starting from it deletes all the vertices in I by ‘time’ T_0 .

Then by union bound

$$(14) \quad \mathbb{P}(E(L)) \leq \sum_{w,I} \mathbb{P}[F(w, I)],$$

where the sum is over all words of length $l \geq L$ and all possible subsets $I \subset [l]$ such that $|I| = l - K$, $w_i = w_{i+1}$ for $i \in I$. Note that such a set I need not exist for all w of length l . We now bound $\mathbb{P}[F(w, I)]$. Let H_w denote the number of atoms (z, v, w) in χ that exists at any time during $[0, S_0]$. By definition of the halving chain in Definition 2 and union bound again, we get

$$(15) \quad \mathbb{P}[F(w, I)] \leq (1 - e^{-T_0})^{l-K} \mathbb{E}(H_w).$$

since for every atom the chance that the vertices corresponding to I will be deleted by ‘time’ T_0 and the expected number of atoms of with word w during $[0, S_0]$ is $\mathbb{E}(H_w)$. Now χ is a PPP whose intensity is given in Definition 3. Let H_w consists of atoms that are born during $(0, S_0]$ and those born before 0 but exists at a positive time. The expected number of the former is $2b(w)/h(w)S_0 \leq 2|w|S_0$. For the latter we observe the following fact that $N_w(0, \cdot)$ is a birth-and-death chain running in stationarity. By Lemma 6 the distribution of the number of atoms of word w that exists at 0 is Poisson($1/h(w)$). Hence, the expected value is at most one. Combining the two, we get

$$(16) \quad \mathbb{E}(H_w) \leq 2S_0|w| + 1.$$

All that remains is to bound the number of possible w and I over which the sum in (14) runs. Now the number of w, I pairs with $|w| = l$ is at most $(2d)^K l^K$. See [JP14, page 1416]. We include the short proof for completeness. For any pair w, I let $u = u_1 u_2 \dots u_K$ be the word of length K obtained from w after deleting the vertices in I . Hence w must necessarily look like

$$(17) \quad \underbrace{u_1 \dots u_1}_{a_1 \text{ times}} \quad \underbrace{u_2 \dots u_2}_{a_2 \text{ times}} \quad \dots \quad \underbrace{u_K \dots u_K}_{a_K \text{ times}}.$$

The total number of choices for u is at most $(2d)^K$ and the total number of choices of $a_1, a_2 \dots a_K$ is at most l^K .

Thus combining this and (16) we get

$$(18) \quad \mathbb{P}[E(L)] \leq (2d)^K \sum_{l=L}^{\infty} (2S_0 l + 1) l^K (1 - e^{-T_0})^{l-K}.$$

The right side of the above bound is summable. Hence, one can find L large enough such that it is smaller than any $\epsilon > 0$.

Outside this event of probability at most ϵ , no atom of word length more than L contributes to $N_w(t, s)$, $(t, s, w) \in R$. This proves (13). The almost sure finiteness follows immediately since $\sup_{(t,s,w) \in R} N_w^{(L)}(t, s)$ is obviously finite for every L . \square

Definition 5 (The doubling chain). The doubling chain is a time homogenous Markov process on the state space \mathcal{W}' with the following transition kernel. Let $u \in \mathcal{W}_k/D_{2k}$ and $w \in \mathcal{W}_{k+1}/D_{2k+2}$ be two words such that w can be obtained from u by doubling a different letters. Then, the chain jumps from u to w at rate a .

An easier description of doubling chain is to attach i.i.d. exponential one random clocks to every letter of the word. Whenever a clock rings, the corresponding letter doubles. In particular, if Y is a doubling chain, then $|Y|$ is a Yule process.

The next lemma shows how the doubling chain can be thought of as the time-reversal of the halving chain.

Lemma 8. *Consider the limiting field in Definition 3. Fix any s . The process $(N_w(t, s), t \in \mathbb{R}, w \in \mathcal{W}')$ is a time homogeneous Markov process with respect to the natural filtration running in stationarity. Looked backwards in ‘time’, the process $(N_w(-t, s), t \in \mathbb{R}, w \in \mathcal{W}')$ is again a time homogenous Markov process running in stationarity. Forward in ‘time’, the process counts existing atoms of the PPP χ performing independent doubling chains while new atoms get born independently. Backward in ‘time’, individual atoms perform independent halving chains. In particular, for any $t \in \mathbb{R}$, the distribution of the vector $(N_w(t, s), w \in \mathcal{W}')$ is always the product of $\text{Poisson}(1/h(w))$, $w \in \mathcal{W}'$.*

Proof. This is essentially the equivalence of [JP14, Lemma 12] and [JP14, Lemma 13]. The argument has been shown for words of bounded size (the bound is called L in those Lemmas). The current lemma follows by an easy extension to $L = \infty$. \square

Recall the notion of (weak) duality for Markov processes from [GS84, Part III].

Lemma 9. *Let $(Q_u, u \geq 0)$ denote the sub-Markovian transition operator for the halving chain on \mathcal{W}' , and let $(\hat{Q}_u, u \geq 0)$ denote the Markovian transition operator of the doubling chain. Consider the measure μ_h on \mathcal{W}' such that $\mu_h(\{w\}) = 1/h(w)$ for all $w \in \mathcal{W}'$. Then Q and \hat{Q} are dual with respect to μ_h . In other words, suppose $q_u(w, \bar{w})$ and $\hat{q}_u(\bar{w}, w)$ are the transition probabilities corresponding to Q_u and \hat{Q}_u , respectively. Then, for any two non-negative functions f, g on \mathcal{W}' , we have*

$$(19) \quad \sum_{\bar{w} \in \mathcal{W}'} \frac{f(\bar{w})}{h(\bar{w})} \sum_{w \in \mathcal{W}'} g(w) \hat{q}_u(\bar{w}, w) = \sum_{w \in \mathcal{W}'} \frac{g(w)}{h(w)} \sum_{\bar{w} \in \mathcal{W}'} f(\bar{w}) q_u(w, \bar{w}).$$

Proof. This follows from Lemma 8. The idea is to consider our counting processes $(N_w(\cdot, 0), w \in \mathcal{W}')$ during any interval of dimension of length u . Let us call this interval $[0, u]$, increasing in dimension from 0 to u . Then atoms exist as a PPP on \mathcal{W}' with intensity μ_h at ‘time’ zero. From every atom we run an independent doubling chain till ‘time’ u . During interval $(0, u)$ new atoms arrive at a certain rate and we start independent doubling chains, all stopped eventually at ‘time’ u . Then, the distribution of counts of various words at ‘time’ u is again a PPP with intensity μ_h . By Lemma 8 the paths of atoms backward in ‘time’ is exactly the halving chain.

Now, consider a pair of words (\bar{w}, w) as an atom itself, representing an atom existing at \bar{w} at ‘time’ zero that moves to w at time u . If there is an atom w at ‘time’ u which has no pre-image at ‘time’ 0, we denote it by (Δ, w) , where Δ is the cemetery of the halving chain. Extend the function f to $\mathcal{W}' \cup \{\Delta\}$ by taking $f(\Delta) = 0$. Define a function F on $\mathcal{W}' \cup \{\Delta\} \times \mathcal{W}'$ as $F(\alpha, \gamma) = f(\alpha)g(\gamma)$. Then, we can count the expected value of the sum of F applied to every such atomic pairs in two ways: one, forward in ‘time’, and the other, backward in ‘time’. The two sides must coincide, and this proves (19). \square

Remark 1. Equation (19) is a generalization of [JP14, Lemma 9] which can be recovered by taking indicator functions $f = \delta_u$, $g = \delta_w$ for any pair of words u, w and then taking derivative with respect to t at $t = 0$.

The following proposition describes the evolution of limiting cycle counts in time at any dimension. Consider the PPP χ described in Definition 3. Fix any $t \leq 0$. Given any atom (z, v, w) of χ , consider the independent halving chain that starts from that atom. Suppose the state of that halving chain at ‘time’ $-t$ is \bar{w} . Then, extend the atom (z, v, w) to (z, v, w, \bar{w}) , and consider the point process $\chi(t) := \sum \delta_{(z, v, \bar{w})}$ by dropping the original word w . Clearly, $(N_w(t, \cdot), w \in \mathcal{W}')$ is a function of $\chi(t)$.

Proposition 10. *The point process $\chi(t)$ is a PPP on $(-\infty, \infty) \times [0, \infty) \times \mathcal{W}'$ with an intensity measure that has a density with respect to the product of the Lebesgue measure on $(-\infty, \infty) \times [0, \infty)$, and the counting measure on \mathcal{W}' . At a section $\{z\} \times [v, \infty) \times \{w\}$ on the state space, the rate is given by*

$$(20) \quad r_t(v, \bar{w}) = \frac{2}{h(\bar{w})} \mathbb{E}_{\bar{w}} \left(b(Y_t) e^{-2b(Y_t)v} \right),$$

where Y_t is a doubling chain at ‘time’ $-t$ and $\mathbb{E}_{\bar{w}}$ denotes expectation with starting state being \bar{w} . In particular, the rate is stationary in the time-coordinate z . The birth rate of atoms with word \bar{w} at any time is exactly

$$(21) \quad r_t(\bar{w}) := r_t(0, \bar{w}) = \frac{2}{h(\bar{w})} (b(\bar{w}) - |w| + e^{-t} |w|), \quad t \leq 0.$$

Proof of Proposition 10. The fact that $\chi(t)$ is a PPP is a consequence of Poisson thinning. Pick a word \bar{w} . The word of every atom $(z, v, w) \in \chi$ has a certain probability of producing $(z, v, \bar{w}) \in \chi(t)$ independent of every other atom. This proves independent Poisson counts over disjoint rectangles, and hence the claim.

We will be done once we compute the intensity measure of this PPP. It is obvious from the structure of χ that the intensity measure is translation invariant in time. Hence, we can restrict ourselves to computing rates at time 0.

Let us first evaluate the birth rates of atoms with words \bar{w} at dimension t . By time stationarity, this is function of (t, \bar{w}) , which we will refer to as $r_t(\bar{w})$. It follows from the PPP structure that

$$r_t(\bar{w}) = \sum_{w \in \mathcal{W}'} \frac{2b(w)}{h(w)} q_{-t}(w, \bar{w}),$$

where q_{-t} , as in Lemma 9 is the transition density of the halving chain.

This allows us to express $r_t(\bar{w})$ as the right side of (19) by taking $f = \delta_{\bar{w}}$, the indicator of the word \bar{w} , and $g(w) = 2b(w)$. Thus, from the left side of (19) we get

$$r_t(\bar{w}) = \frac{2}{h(\bar{w})} \mathbb{E}_{\bar{w}} (b(Y_t)),$$

where Y is a doubling chain and Y_t refers to state of the chain at ‘time’ $-t$.

But, every doubling increases both the size of the word and the value of b by exactly one. Thus $b(Y_t) = b(\bar{w}) + |Y_t| - |\bar{w}|$. However, as discussed above, $|Y|$ is a Yule process, and hence $\mathbb{E}(|Y_t|) = |w| e^{-t}$. By substituting above, we get

$$(22) \quad r_t(\bar{w}) = \frac{2}{h(\bar{w})} (b(\bar{w}) - |w| + e^{-t} |w|).$$

Let us now compute the joint intensity of birth and lifetimes of atoms with word \bar{w} at dimension t . Suppose that a word w at dimension 0 gets reduced to word \bar{w} at dimension t . The corresponding lifetime still remains exponential with rate $2b(w)$. In particular, we see that the lifetime of atoms at any dimension $t < 0$ is not exponential, unlike the case at dimension zero.

In fact, the intensity measure of the pair birth and lifetime of an atom with word \bar{w} can be easily seen from Poisson counting. Fix word \bar{w} and $v > 0$. Let, as in the statement, $r_t(v, \bar{w})$ be the rate at which atoms with word \bar{w} and lifetime in $[v, \infty)$ are getting created at dimension $-t$ and time zero. Then, as above,

$$r_t(v, \bar{w}) = \sum_{w \in \mathcal{W}'} \frac{2b(w)}{h(w)} e^{-2b(w)v} q_{-t}(w, \bar{w}).$$

This is the right side of (19) when we take $f = \delta_{\bar{w}}$ and $g(w) = 2b(w)e^{-2b(w)v}$.

Therefore, by Lemma 9, if Y_t is the state of a doubling chain at ‘time’ $-t$, then

$$r_t(v, \bar{w}) = \frac{2}{h(\bar{w})} \mathbb{E}_{\bar{w}} \left(b(Y_t) e^{-2b(Y_t)v} \right).$$

The above can again be computed explicitly in terms of Yule processes, but this is unnecessary for our analysis. \square

3.1. The topology of convergence. We have stochastic processes with multidimensional parameters. The topology of weak convergence that we choose to work with is described in [Neu71]. This is a generalization of the usual Skorokhod space of RCLL paths. For the benefit of the reader we give a short informal introduction. For more details, please consult [Neu71].

Let I_1, I_2 be two bounded and closed intervals in \mathbb{R} . The space of surfaces that we will consider will be denoted by $D(I_1 \times I_2)$. By shifting and scaling we can assume that $I_1 = I_2 = [0, 1]$, and we will denote the corresponding space by D .

To define elements in D , we define *quadrants*. Fix any $(t, s) \in [0, 1] \times [0, 1]$. Then the four quadrants are the four open subsets of $[0, 1] \times [0, 1]$ given by removing the axes passing through (t, s) :

$$\{[0, t) \times [0, s), [0, t) \times (s, 1], (t, 1] \times [0, s), (t, 1] \times (s, 1]\}.$$

Some of these are empty when (t, s) lies on the boundary of $[0, 1] \times [0, 1]$.

We now generalize the RCLL property. For a function f on $[0, 1] \times [0, 1]$, we say that its quadrant limits exist at (t, s) if, for every non-empty quadrant Q at (t, s) , and any sequence of points $\{(t_n, s_n), n \in \mathbb{N}\} \subseteq Q$ such that $\lim_n (t_n, s_n) = (t, s)$, the quantity $\lim_n f(t_n, s_n)$ exists.

This does not say anything about the value of the function at (t, s) . For every $(t, s) \in [0, 1) \times [0, 1)$, consider the special up-right quadrant $(t, 1] \times (s, 1]$. If either s or t is 1, the special quadrant is given by considering the interval $[0, 1)$ for that coordinate (instead of the empty set $(1, 1]$). We say that the function f is continuous is continuous from above at (t, s) if the quadrant limit in the special quadrant is equal to the value $f(t, s)$.

We now define the space D to be the space of all real valued functions on $[0, 1] \times [0, 1]$ which have quadrant limits and is continuous from above at every point. It follows that such functions are bounded, are RCLL in the traditional sense along every line parallel to the axes, and have at most countably many jumps.

The Skorokhod topology on D is an extension of the usual Skorokhod topology [Neu71, Section 2]. As usual, by defining a proper metric, the space can be turned

to a complete separable metric space over which we can define weak convergence. We will provide citations as needed later.

For the rest of the section we will need the product Skorokhod topology on $D^{\mathcal{W}'}$ or $D^{\mathbb{N}}$. The notion of convergence in this product is pointwise convergence of every coordinate. In particular, we will use the fact that marginal tightness along every coordinate implies joint tightness. This is a consequence of Tychonoff's theorem.

3.2. Gaussian limits and covariance computation. Recall the cycle counting field $(N_k(t, s))$ from (5). We now compute the asymptotic covariance of any pair of elements $(N_j(t_1, s_1), N_k(t_2, s_2))$, $t_1, t_2 \in (-\infty, 0]$, $s_1, s_2 \in [0, \infty)$, from this field, where the ordering of the pair implies $s_1 \leq s_2$. This constitutes part of Theorem 2. The next lemma computes this covariance for every d .

Lemma 11. *We have the following cases:*

- (i) *If $j \neq k$ and $t_1 = t_2$, then $\text{Cov}(N_j(t_1, s_1), N_k(t_2, s_2)) = 0$.*
- (ii) *If $j \neq k$ and $t_1 \neq t_2$, then*

$$0 \leq \text{Cov}(N_j(t_1, s_1), N_k(t_2, s_2)) \leq c_0(2d-1)^{j \wedge k},$$

where the constant c_0 depends on j, k, t_1, t_2, s_1, s_2 , but not on d .

- (iii) *If $j = k$ and $t_1 = t_2 = t$, then*

$$\text{Cov}(N_j(t, s_1), N_j(t, s_2)) = \frac{a(d, j)}{2j} \mathbb{E} \left[\frac{1}{h(w)} e^{-2b(Y_t)s} \right],$$

where \mathbb{E} represents the joint law of a word w chosen uniformly at random from \mathcal{W}_j/D_{2j} and Y_t is the state of a doubling chain, starting from w , at 'time' $-t$.

- (iv) *Finally, if $j = k$ and $t_1 \neq t_2$, then*

$$\text{Cov}(N_j(t_1, s_1), N_j(t_2, s_2)) = \frac{a(d, j)}{2j} e^{-j|t_1 - t_2|} \mathbb{E} \left[\frac{1}{h(w)} e^{-2b(Y_t)s} \right], \quad t = \max(t_1, t_2),$$

where \mathbb{E} is the probability measure described in (iii).

Proof of Lemma 11. It follows from time stationarity, that, without loss of generality, we can replace the pair (s_1, s_2) by $(0, s)$, where $s = s_2 - s_1$.

Consider case (i). Let $t_1 = t_2 = t$. Suppose $j \neq k$. then N_j and N_k counts atoms of the PPP $\chi(t)$ over disjoint collections of words. Therefore, they are independent and has zero covariance.

Consider case (ii). We compute $\text{Cov}(N_j(t_1, 0), N_k(t_2, s))$. There are two sub-cases: either $t_1 < t_2$ or $t_1 > t_2$. Since the PPP χ is time-reversible, these two cases are symmetric. Hence, without loss of generality, we consider the case of $t_1 < t_2$.

Consider $\chi(t_2)$. The atoms counted in $N_j(t_1, 0)$ are obtained as a Poisson thinning of atoms of $\chi(t_2)$ that exist at time zero. The atoms counted in $N_j(t_2, s)$ consists of two independent collections: those that exist at both time zero and time s , and those that were born after time zero but exist at time s . Thus

$$0 \leq \text{Cov}(N_j(t_1, 0), N_k(t_2, s)) \leq \text{Cov}(N_j(t_1, 0), N_k(t_2, 0)).$$

The last covariance, computed in [JP14, Corollary 17], produces the bound.

Now, consider case (iii). As before, we compute $\text{Cov}(N_j(t, 0), N_j(t, s))$, where $s = s_2 - s_1 \geq 0$. Consider atoms of $\chi(t)$. The atoms counted in $N_j(t, 0)$ can be classified in two groups: either existing simultaneously at both time 0 and s , or not. The same holds for atoms counted in $N_j(t, s)$. By Poisson thinning

$$\text{Cov}(N_j(t, 0), N_j(t, s)) = \text{Var}(N_j(t, [0, s])),$$

where $N_j(t, [0, s])$ is the number of atoms of $\chi(t)$ that exist simultaneously at both times 0 and s .

Define a Borel subset $\Gamma \subseteq (-\infty, \infty) \times [0, \infty) \times \mathcal{W}'$ by

$$\Gamma := \{(z, v, w) : z \leq 0, v \geq s - z, |w| = j\}.$$

Then $N_j(t, [0, s])$ is the mass that the Poisson random measure $\chi(t)$ puts on Γ . In particular, it is a Poisson random variable whose expectation and variance are both given by the mass of the intensity measure on Γ .

The variance can now be computed using (21):

$$\begin{aligned} \text{Var}(N_j(t, [0, s])) &= \sum_{w \in \mathcal{W}_j/D_{2j}} \frac{2}{h(w)} \int_{-\infty}^0 \mathbb{E}_w \left[b(Y_t) e^{-2b(Y_t)(s-z)} \right] dz \\ &= \sum_{w \in \mathcal{W}_j/D_{2j}} \frac{2}{h(w)} \mathbb{E}_w \left[b(Y_t) \int_{-\infty}^0 e^{-2b(Y_t)(s-z)} dz \right], \quad \text{By Fubini-Tonelli,} \\ (23) \quad &= \sum_{w \in \mathcal{W}_j/D_{2j}} \frac{1}{h(w)} \mathbb{E}_w \left[2b(Y_t) \int_s^\infty e^{-2b(Y_t)u} du \right] \\ &= \sum_{w \in \mathcal{W}_j/D_{2j}} \frac{1}{h(w)} \mathbb{E}_w \left(e^{-2b(Y_t)s} \right). \end{aligned}$$

The claimed statement follows since the number of elements in \mathcal{W}_j/D_{2j} is $a(d, j)/2j$.

Finally, we consider case (iv). By symmetry, as in case (ii), we can assume $t_1 < t_2$. Again, by Poisson thinning, we can decompose both $N_j(t_1, 0)$ and $N_j(t_1, s)$ as a sum of several independent Poisson random variables with exactly one common class counted in both of them. This is the count of all atoms in $\chi(t_2)$ that exist simultaneously at both times 0 and s at dimension t_2 , and moreover, the halving chains starting from those atoms do not jump during ‘time’ $[0, t_2 - t_1]$. This is because other atoms counted in $N_j(t_1, 0)$ either do not exist at time s at dimension t_2 , or must have descended from words of a bigger size at dimension t_2 . Both collections are independent from atoms counted in $N_j(t_2, s)$.

Thus

$$\text{Cov}(N_j(t_1, 0), N_j(t_2, s)) = \text{Var}(B),$$

where the random variable B is Binomial, given $N_j(t_2, [0, s])$, with parameters $N_j(t_2, [0, s])$ and \hat{p} . Here \hat{p} represents the probability that a halving chain starting from a word w , with $|w| = j$, does not jump during ‘time’ $t_2 - t_1$. By definition, $\hat{p} = e^{-(t_2 - t_1)j}$. Therefore, B is Poisson with parameter $\hat{p} \mathbb{E}(N_j(t_2, [0, s]))$. Substituting the values from case (iii) computes the expression for its variance. \square

Lemma 12. *Fix $j \in \mathbb{N}$. Let W be a uniformly picked word in \mathcal{W}_j/D_{2j} . Then, as d tends to infinity, the asymptotic law of $|W| - b(W)$ is that of the number of sign changes along a j -cycle if we attach random i.i.d. ± 1 at every vertex. The asymptotic law of $h(W)$ is the delta mass at one. In particular $(b(W), h(W))$ are asymptotically independent.*

Proof of Lemma 12. One can imagine W as a random pick from all possible cyclic words (up to equivalent classes) of length j , conditioned on being cyclically irreducible. A random pick from all possible cycles can be generated by picking i.i.d. elements from the $2d$ collection of letters $\{\pi_i, \pi_i^{-1}, i \in [d]\}$ at every edge of a j -cycle. The expected number of occurrences of successive letters $\pi_i \pi_i^{-1}$ or $\pi_i^{-1} \pi_i$ is

$j/2d$. Therefore, by Markov's bound, the probability that such a cycle is *not* cyclically irreducible is vanishing as d tends to infinity. Now, when each letter is picked independently, their signs are distributed as independent coin tosses. Hence the asymptotic law of $j - b(W)$. The asymptotic law of $h(W)$ follows by counting primitives. \square

Proof of Theorem 2. By usual CLT for poisson variables and (7) it is clear that there is finite-dimensional convergence of $(X_j(t, s), (t, s, j) \in (-\infty, 0) \times (0, \infty) \times \mathbb{N})$ to a centered Gaussian field $(U_j(t, s), (t, s, j) \in (-\infty, 0) \times (0, \infty) \times \mathbb{N})$. Let us first argue that the collection of Gaussian random surfaces has the stated covariance structure.

It follows immediately from (i), (ii) in Lemma 11 that, if $j \neq k$, then

$$\text{Cov}(U_j(t_1, s_1), U_k(t_1, s_2)) = 4jk \lim_{d \rightarrow \infty} (2d-1)^{(j+k)/2} \text{Cov}(N_j(t_1, s_1), N_k(t_2, s_2)) = 0.$$

Extending the argument to linear combinations of U_j and U_k at different points in dimension and time proves that the entire fields U_j and U_k are independent.

Now take $j = k$ and $t_1 = t_2 = t$. Assume as before, $s_1 = 0$ and $s_2 = s \geq 0$. Recall from (6)

$$\lim_{d \rightarrow \infty} (2d-1)^{-j} a(d, j) = 1, \quad j \in \mathbb{N}.$$

Then, from Lemma 11, it follows that

$$\text{Cov}(U_j(t, 0), U_j(t, s)) = \frac{4j^2}{2j} \lim_{d \rightarrow \infty} \mathbb{E} \left[\frac{1}{h(w)} e^{-2b(Y_t)s} \right] = 2j \lim_{d \rightarrow \infty} \mathbb{E} \left[e^{-2b(Y_t)s} \right].$$

The final equality is due to Lemma 12.

Now, fix d , and consider $b(Y_t)$. As in the derivation of (22), we can write

$$b(Y_t) = b(W) + |Y_t| - j,$$

where W is a randomly chosen word of length j and $|Y_t|$ is a Yule process starting at j , independent of W . Thus

$$\mathbb{E} \left[e^{-2b(Y_t)s} \right] = e^{2sj} \mathbb{E} \left(e^{-2sb(W)} \right) \mathbb{E} \left(e^{-2s|Y_t|} \right).$$

Let $\xi(t)$ be a Yule process at time $-t$ starting with $\xi_0 = 1$. Then, we know that $|Y_t|$ has the same law as the sum of j many independent copies of $\xi(t)$. Therefore, $\mathbb{E} \left(e^{-2s|Y_t|} \right) = \left[\mathbb{E} \left(e^{-2s\xi(t)} \right) \right]^j$.

Combining all the pieces, we get

$$\begin{aligned} \text{Cov}(U_j(t, 0), U_j(t, s)) &= 2j \left[\mathbb{E} \left(e^{-2s\xi(t)} \right) \right]^j \lim_{d \rightarrow \infty} \mathbb{E} \left(e^{2s(j-b(W))} \right) \\ &= 2j \left[\mathbb{E} \left(e^{-2s\xi(t)} \right) \right]^j \mathbb{E} \left(e^{2s\tau} \right), \end{aligned}$$

where τ is the asymptotic law of $|W| - b(W)$ as described in Lemma 12. The general case follows along similar lines from Lemma 11 (iv).

We now fix $k \in \mathbb{N}$ and argue marginal tightness of the field $(X_k(t, s), t \leq 0, s \geq 0)$. We fix some rectangle $[-T_0, 0] \times [0, S_0]$. The argument is similar to the case of $T_0 = 1, S_0 = 1$, which is what we assume for the rest of the proof.

Consider the PPP χ from Definition 3 and the independent halving chains starting from its atoms. Let Q_w denote the law over the Skorokhod space $D[0, 1]$ (for more details about the Skorokhod space see [Bil99]) of the halving chain starting

at word w . Then, one can think of the collection of atoms in χ and the halving chains as not separate entities but points of a PPP $\tilde{\chi}$ on

$$(-\infty, \infty) \times [0, \infty) \times D[0, 1]$$

with an intensity measure that is described below. Consider an atom (z, v, x_w) , where X_w is a path of a halving chain starting at word w . It occurs at a rate that is the product of the rate of occurrence of (z, v, w) in χ and $Q_w(dx_w)$.

Now consider $N_k(t, s)$. One can write it as the sum of coordinatewise monotone processes in the following way. Let

$$H_1(t, s) = \sum_{(z, v, x_w) \in \tilde{\chi}} 1 \{z \leq s, z + v \geq 0, |w| \geq k, |x_w(-u)| = k, \text{ for some } u \in [t, 0]\},$$

$$H_2(t, s) = \sum_{(z, v, x_w) \in \tilde{\chi}} 1 \{z \leq s, z + v \geq 0, |w| \geq k, |x_w(-u)| = k - 1, \text{ for some } u \in [t, 0]\}.$$

In other words, $H_1(t, s)$ is the cumulative count of all atoms of $\tilde{\chi}$ that exist at some point during interval $[0, s]$ with word of size at least k such that the halving chain from that word has size exactly k at some dimension in $[t, 0]$. Clearly $H_1(t, s)$ is increasing along s and decreasing along t (since $t < 0$), and is distributed as Poisson for every fixed (t, s) . Similarly, H_2 counts those among H_1 that have jumped to a size below k . Thus, $H_1(t, s) - H_2(t, s)$ counts the number of atoms that are born or exists at some point in $[0, s]$ and are of size exactly k at dimension t .

Similarly, let

$$H_3(t, s) = \sum_{(z, v, x_w) \in \tilde{\chi}} 1 \{0 \leq z + v \leq s, |w| \geq k, |x_w(-u)| = k \text{ for some } u \in [t, 0]\}$$

$$H_4(t, s) = \sum_{(z, v, x_w) \in \tilde{\chi}} 1 \{0 \leq z + v \leq s, |w| \geq k, |x_w(-u)| = k - 1 \text{ for some } u \in [t, 0]\}.$$

This is, $H_3(t, s)$ the cumulative count of all atoms that exist during $[0, s]$ and dies before time s , and is of size k at some dimension larger than t . This is again increasing along s and decreasing along t and is marginally Poisson. A similar interpretation holds for H_4 . Thus, $H_3(t, s) - H_4(t, s)$ counts atoms of size k at dimension t that have died during time $[0, s]$.

Most importantly,

$$(24) \quad \begin{aligned} N_k(t, s) &= H_1(t, s) - H_2(t, s) - (H_3(t, s) - H_4(t, s)), \\ E(N_k(t, s)) &= E(H_1(t, s)) - E(H_2(t, s)) - E(H_3(t, s)) + E(H_4(t, s)). \end{aligned}$$

We now claim that it is enough to show the tightness of each

$$S_i \triangleq (2d - 1)^{-1/2} (H_i - EH_i)$$

in the D topology. The reason is the Continuous Mapping Theorem. The limit of each S_i will turn out to be a continuous Gaussian surface. Thus, under the product topology, the vector $(S_i, i \in [4])$ has an almost sure continuous limit. Also it is not hard to see that convergence to a continuous surface in the D topology in [Neu71] is the same as convergence in the uniform topology when restricted to continuous surfaces. This follows in the same way as for the classical Skorokhod topology. Thus, by Continuous Mapping Theorem, we can exchange the operations of limit and sums in $X_k = S_1 - S_2 - S_3 + S_4$. This proves that X_k is tight.

The proof of tightness of every S_i is similar. So we only explain in detail the case of S_1 . The counting process H_1 itself can be decomposed in three parts. First, separately count of atoms that exist at time 0, and those that were born after time zero. Second, among those born after time zero, count separately those which are born with words of size k and those with size larger than k .

That is, define

$$\begin{aligned} H_1^{(0)}(t) &= \sum_{(z,v,x_w) \in \tilde{\chi}} 1 \{z \leq 0, z+v \geq 0, |w| \geq k, |x_w(-u)| = k, \text{ for some } u \in [t, 0]\}, \\ H_1^{(1)}(t, s) &= \sum_{(z,v,x_w) \in \tilde{\chi}} 1 \{0 < z \leq s, z+v \geq 0, |w| = k\}, \\ H_1^{(2)}(t, s) &= \sum_{(z,v,x_w) \in \tilde{\chi}} 1 \{0 < z \leq s, z+v \geq 0, |w| > k, |x_w(-u)| = k, \text{ for some } u \in [t, 0]\}. \end{aligned}$$

The corresponding centered and scaled processes, $S_1^{(0)}$, $S_1^{(1)}$, and $S_1^{(2)}$, can be similarly defined.

Notice that $H_1^{(0)}(t)$ does not depend on s . It can be extended to a surface by defining $H_1^{(0)}(t, s) \equiv H_1^{(0)}(t)$. It follows from [Neu71, Section 2] that the D topology restricted to surfaces that are constant in the time axis is the usual Skorokhod topology for the process restricted to the dimension axis. It is not hard to see that $(S_1^{(0)}(t), -1 \leq t \leq 0)$ is a centered and scaled Poisson process and hence converges to Brownian motion in $D[0, 1]$ and therefore has a continuous limit. For details see [JP14, page 1425] and [Joh14]. Thus it has a continuous limit in D for the entire surface.

Similarly $H_1^{(1)}(s) = H_1^{(1)}(t, s)$ does not depend on t and a similar argument like above shows continuous limit for $(H_1^{(1)}(s), 0 \leq s \leq 1)$.

Finally, for $S_1^{(2)}$ we relate this problem to the empirical process X_n^F considered in [Neu71, Section 5]. The idea is the following, for every atom (z, v, w) of $\tilde{\chi}$ that is counted in $H_1^{(2)}$, consider the point (t, z) on $[-1, 0] \times [0, 1]$, where $-t$ is the first ‘time’ the chain x_w hits a word of size k . The collection of points thus created is a PPP which can be described as the empirical process of i.i.d. many points in the following way. Condition on the number of points in the rectangle $[-1, 0] \times [0, 1]$. There are finitely many points which, by the PPP structure, are distributed independently and identically on the rectangle. A typical point (T, Z) has independent coordinates: Z is distributed uniformly over $[0, 1]$, while T has a continuous distribution of the first hitting time of size k of a pure death chain, conditioned to be less than $-t$. Thus, conditioned on the number of points, our counting process $S_1^{(2)}$ is basically the process $X_n^F(\cdot)$ in [Neu71, eqn. (4.2)] (see the remark preceding it). Since the number of points is Poisson with a mean going to infinity with d , a standard de-Poissonization argument extends the convergence argument in [Neu71] to our case. This proves a continuous limit for $H_1^{(2)}$.

Combining all three pieces and invoking a similar application of Continuous Mapping Theorem as before, we get a continuous Gaussian limit for S_1 . Combining similar statements for S_2, S_3 , and S_4 , we get tightness for the surface X_k .

Since the limit of each S_i is a continuous surface, the limit of X_k must also be continuous. Since marginal tightness implies joint tightness in the product topology, this completes the proof of Theorem 2. \square

Remark 2. Since the marginal distribution of the Yule process $\xi(t)$ is geometric with parameter e^t , each of the covariances in Theorem 2 can be computed explicitly. But again, this is unnecessary for our analysis.

For the covariance in the stationary regime of Theorem 3, we recall the following standard limit theorem for Yule processes.

Lemma 13. *Let $\xi(\theta)$, $\theta \geq 0$ be a Yule process such that $\xi(0) = 1$. Then $e^{-\theta}\xi(\theta)$ converges in law to an exponential one random variable. Thus, for any $v \geq 0$, we have*

$$(25) \quad \lim_{\theta \rightarrow \infty} \mathbb{E} \exp(-ve^{-\theta}\xi(\theta)) = \frac{1}{1+v}.$$

Proof. The first claim is classical. The exact distribution at time θ is a geometric with mean e^θ . See [KT75, pg 122]. Convergence to an exponential one now follows from that. Equation (25) follows from the stated weak convergence since e^{-vx} is a continuous bounded function for $x \geq 0$. \square

Proof of Theorem 3. Let us show that the covariance converge to the stated limit. Pick $u_1, u_2, v_1, v_2 \geq 0$, $j \in \mathbb{N}$ and T_0 large enough. Define

$$t_1 = -T_0 + u_1, \quad t_2 = -T_0 + u_2, \quad s_1 = \frac{1}{2}v_1e^{-T_0}, \quad s_2 = \frac{1}{2}v_2e^{-T_0}.$$

Let $s = |s_1 - s_2|$. Then, obviously $\lim_{T_0 \rightarrow \infty} \mathbb{E}(e^{2s\tau}) = 1$. Let $\theta = -\max(t_1, t_2) = T_0 - u_1 \vee u_2$. Then, by Lemma 13,

$$\begin{aligned} \lim_{T_0 \rightarrow \infty} \mathbb{E}(e^{-2s\xi(\theta)}) &= \lim_{T_0 \rightarrow \infty} \mathbb{E}(\exp(-|v_1 - v_2|e^{-T_0}\xi(T_0 - u_1 \vee u_2))) \\ &= \frac{1}{1 + |v_1 - v_2|e^{-u_1 \vee u_2}}. \end{aligned}$$

Combining all the pieces from Theorem 2, gives us the correct covariance.

We only need to argue tightness of each U_j as $T_0 \rightarrow \infty$. There are many ways to argue weak convergence of Gaussian surfaces. We choose to use [DZ08, Theorem 1]. The topologies allowed in [DZ08] includes that of [Neu71] (over continuous surfaces they are all uniform convergence). See the discussion at the beginning of Section 2 in [DZ08] and the discussion following Theorem 3.1 in [Neu71].

Take any sequence of T_0 's growing to infinity. In the notation of [DZ08, Theorem 1], we have $d = 2$, and take $\alpha = 6$, $\beta = 3$, and $a_{T_0} = 1/T_0$. It follows from the covariance convergence that for any (t_1, s_1) and (t_2, s_2) , the random variable $U_j(t_1, s_1) - U_j(t_2, s_2)$ is Gaussian with mean zero and a variance that is of the order

$$(26) \quad O(\max(|t_1 - t_2|, |s_1 - s_2|)),$$

where the O can be taken not to depend on T_0 . This shows

$$(27) \quad \mathbb{E}|U_j(t_1, s_1) - U_j(t_2, s_2)|^6 = O(|t_1 - t_2|^3 + |s_1 - s_2|^3)$$

and hence condition (1) in [DZ08, Theorem 1]. Condition (2) follows since the Gaussian field U_j is uniformly Hölder continuous as $T_0 \rightarrow \infty$. This is a consequence of the Kolmogorov-Čentsov Theorem and the uniform bound (27). For details, see

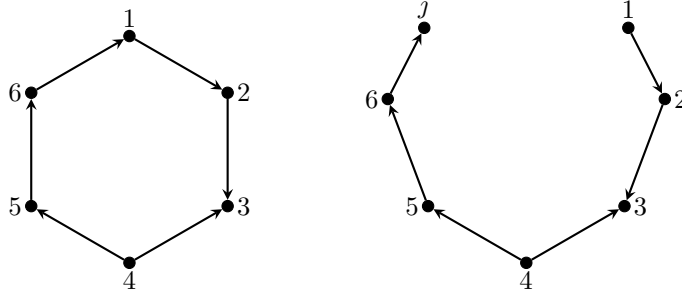


FIGURE 2. A cycle that vanishes due to the action of the transposition $(1, j)$, $j > 6$, that gets multiplied on the left.

[KS91, page 53]. The uniform moment bound for the Gaussian random variable $U_j(0, 0)$ follows from the fact that its variance has a limit as $T_0 \rightarrow \infty$. These verify all the conditions for [DZ08, Theorem 1]; in particular, we obtain that the limiting Gaussian field G_j is continuous almost surely. This completes the proof. \square

4. WEAK CONVERGENCE OF CYCLE COUNTS

4.1. Heuristic arguments on the limit. In this subsection we give heuristic arguments to justify the form of the limiting field. Suppose $M_T = n$, i.e., the graph $G(T, 0)$ has n vertices, where n is very large. Consider a cycle with word w . For our purpose it suffices to consider that the vertex labels and edge directions are given while the edge labels are omitted. See Figure 2 which depicts one such. For any word w , recall that $C_w(T, s)$ is the count of the number of cycles with word w at dimension T and time s .

Consider the possible ways this cycle can get modified under the action of a transposition. Consider some $j > 6$. The transposition $(1, j)$ opens the cycle up (as in Figure 2). The same happens when 1 is replaced by any other vertex of the cycle such that the direction of the edges is the same on both sides of it. The number of such vertices is $b(w)$. Other possibilities are multiplying with $(3, j)$ which replaces 3 by j but does not effect the count $C_w(T, \cdot)$. Nothing changes at all when multiplied with $(4, j)$. There are other possibilities. For example, we could choose a transposition (i, j) where both $i, j \in [6]$. However, these events are of negligible probability for large T . Hence, the approximate rate at which this cycle ceases to exist is $2b(w)$ (recall definition from Definition 1), at which point the count $C_w(T, \cdot)$ decreases by one.

The other possibility is the appearance of a new cycle of word w . The easiest way to calculate the rate is to appeal to stationarity. From [JP14, Thm 14, Cor 15] we know that the law of $C_w(T, 0)$ is approximately Poisson with mean $1/h(w)$. By exchangeability of vertex labels, the same law is true for any $C_w(T, s)$. Because the underlying graph is large and the number of cycles is roughly of constant order it is not too hard to imagine that the rate at which new cycles form should roughly stay the same along time. Thus by the above discussion it follows that the law of $C_w(T, \cdot)$ for large T is a birth-and-death Markov chain. Since cycles disappear at rate $2b(w)$, they constant birth rate must be $2b(w)/h(w)$ in order to keep the given Poisson distribution invariant.

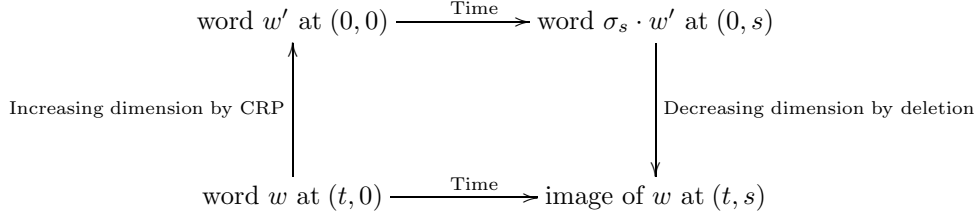


FIGURE 3. Joint dynamics of a cycle at two times.

If cycles do not share vertices it seems reasonable that individual cycles get born and die independently of one another. Hence, the joint law of the process $(C_w(T, s), w \in \mathcal{W}', s \geq 0)$ is approximately given by independent birth-and-death chains where the individual laws are described above.

Another way of expressing this birth-and-death structure is to think of cycles appearing as a Poisson point process on the time axis according to a rate that depends on the word. With every atom that represents a cycle being born, we attach the length of time the cycle survives. These lifetimes are roughly independent exponentials, and this gives us the limiting PPP χ in Definition 3.

We now track these cycles backward in dimension. Suppose a cycle with word w' exists at dimension T and time 0. At time 0, looked backwards in dimension, the cycle shrinks in length or disappears entirely. The resulting sequence of words follow the halving chain. See the heuristics in [JP14, Section 3.1] and also [JP14, Lemma 13]. Suppose now that this cycle exists simultaneously at two time points (say) 0 and s . The joint law of the CRP backwards at these two time points depends on the order in which we remove the vertices. If we follow our convention outlined in (2) the following convenient feature emerges. Consider the first time it gets halved, say $t_0 < T$, at which point its word becomes w . Then, some letter, say π , of w doubles to give us w' and this is the only difference between the two words. Therefore, on the cycle with word w' , we have a sequence of vertices $i \xrightarrow{\pi} j \xrightarrow{\pi} k$ as three successive vertices with labeled directed edges. Now, as we move time to s , the vertices of this cycle are exactly σ_s applied to the vertices of w' . Now reduce its dimension at time s and track the change in the cycle. Since j is the first vertex to be deleted from the cycle at time 0, by our convention, $\sigma_s(j)$ is the first vertex to be deleted at time s . Thus, the first change to $\sigma_s \cdot w'$ also happens exactly at dimension t_0 when we erase vertex $\sigma_s(j)$ and halve the double letters $\pi \cdot \pi$.

Hence, by looking back at the most recent change at each time, inductively, allows us to describe the joint law of the process of cycles backward in the dimension parameter. Namely, at dimension T , consider any cycle that is born. During the entire time interval of its existence, its path, looked backwards in dimension, is identical up to relabeling of vertices (see Figure 3). At any time point during its existence, this path is a typical path of the halving chain. This produces the limiting cycle counting field described in Theorem 1.

The rest of the section is devoted to the proof of Theorem 1.

4.2. Proof of Theorem 1. Notice that the time parameter for χ ranges over $(-\infty, \infty)$. This is done for a neater description. For the proof below we will work with a restricted version of χ with time varying over $[0, \infty)$.

This is achieved by collecting all atoms of χ that *exists* at time 0 and marking them as points born at time zero. More formally, if $(z, v, w) \in \chi$ such that $z < 0$ and $z + v > 0$, then, we replace this atom by another $(0, v', w)$, where $v' = z + v$. This produces an atomic intensity at time 0. It follows from Lemma 6 and Proposition 10 that for a word w , the number of atoms of word w at time 0 is Poisson with mean $1/h(w)$. By the memoryless property, the remaining lifetimes of atoms remain exponentially distributed. By an abuse of notation, we will continue to call this PPP restricted χ or just χ , in case there is no scope of confusion.

Fix $T > 0$. For every cycle appearing the process $G(T, \cdot)$, consider the triplet (z, v, w) where (i) z is the time when it first appears ($z = 0$ if the cycle exists at time 0), (ii) v is the difference between the time it disappears and z (the lifetime), and (iii) w is the word of the cycle. Construct a point process $\chi(T)$ on $[0, \infty) \times [0, \infty) \times \mathcal{W}'$ as a random measure that counts these atoms (z, v, w) . Hence $\chi(T)$ is similar to χ but for the finite graph process at dimension T . For every $L \in \mathbb{N}$, let $\chi_L(T)$ be the restriction of $\chi(T)$ to atoms whose words are of length at most L . We start by showing that every $\chi_L(T)$ converges to the claimed weak limit χ_L , which is χ restricted to time $[0, \infty)$ and words of length at most L .

Now, for every $T > 0$, consider independent halving chains with initial condition given by atoms of $\chi_L(T)$ as described in Definition 3. There are only finitely many such chains and this operation is well-defined. One can define a cycle counting field $N_w^{(T,L)}(t, s)$ induced by these Markov chains exactly as in (4).

Fix any positive T_0, S_0 . Consider both the count of cycles

$$C_w(T + t, s) \text{ and } N_w^{(T,L)}(t, s), \quad (t, s) \in [-T_0, 0] \times [0, S_0].$$

By trivial modifications at their finitely many jump points, they can be both turned into *primitive functions* in the sense of [Neu71, p. 1288], and, therefore, elements in

$$D_2 \triangleq D^{\mathcal{W}'}([-T_0, 0] \times [0, S_0]).$$

We have the following proposition. The topology of convergence of point processes is the usual one for random Radon measures. See [Res07, Chapter 3].

Proposition 14. *For any $L > 0$, the point process $\chi_L(T)$ converges weakly to the restricted PPP χ_L as T goes to infinity. Moreover, for any $K \in \mathbb{N}$ and any $\epsilon > 0$ one can find an $L \in \mathbb{N}$ such that*

$$(28) \quad \limsup_{T \rightarrow \infty} \mathbb{P} \left(\sup_{(t,s) \in [-T_0, 0] \times [0, S_0], |w| \leq K} |C_w(T + t, s) - N_w^{(T,L)}(t, s)| > 0 \right) < \epsilon.$$

In particular, for large enough T , the total variation distance between

$$(29) \quad C_w(T + t, s) \text{ and } N_w^{(T,L)}(t, s), \quad (t, s) \in [-T_0, 0] \times [0, S_0], \quad w \in \mathcal{W}'_K,$$

is less than ϵ .

4.3. PPP convergence at the front. We start by proving weak convergence of $\chi_K(T)$ to the limiting PPP χ_K at the front for every $K \in \mathbb{N}$.

Proposition 15. *Let \mathbb{P} refer to the law of the transposition Markov chain acting on d uniform random permutations. Then, under \mathbb{P} , for any fixed $K > 0$, the random measure $\chi_K(T)$, converges in law to the restricted PPP χ_K .*

To prove the above proposition we start with the following definitions inspired from [LP10] and earlier work on word maps.

Definition 6. Let w be a word (not an equivalent class). A trail with word $w = w_1 w_2 \dots w_k$ is an edge-labeled directed graph of the form

$$(30) \quad s_0 \xrightarrow{w_1} s_1 \xrightarrow{w_2} s_2 \xrightarrow{w_3} \dots \xrightarrow{w_k} s_k$$

with $s_i \in \{1, \dots, n\}$ and does not have repeated vertices except perhaps at the two ends. The trail is said to be closed if $s_k = s_0$. Clearly, closed trails are cycles with word w . A pre-cycle with word u is a trail that is *not* closed and that can be obtained by multiplying (on left) a closed trail with word u with some transposition σ . Multiplication here means the natural action of the transposition on the edges of the trail. Since multiplication by a transposition is an involution, σ multiplied to the pre-cycle gives us a cycle with word u . This is only possible if the signs of u_1 and u_k are the same, $\sigma = [s_0, s_k]$, where $|u| = k$. Let S_u denote the set of pre-cycles with word u . Two pre-cycles with words u_1 and u_2 are called equivalent if $u_1^{-1} = u_2$ and the sequence of vertices in the first pre-cycle is the reverse of that in the second.

Consider some $T > 0$ and let M_T be the number of vertices of $G(T, 0)$. Condition on $M_T = n$. For the proofs in this subsection, we will send n to infinity instead of T . This is equivalent by the well-known fact. Assume that at dimension zero, the permutations have exactly one label, 1. Then M_T is the state of a Yule process at time T starting with one individual. We know from Lemma 13 that the weak limit $\lim_{T \rightarrow \infty} e^{-T} M_T$ is exponential with mean one. Thus, sending T to infinity is equivalent to sending M_T to infinity.

We now define an appropriate filtration. Let \mathcal{G}_0 denote the σ -algebra generated by the d many permutations $\{\pi_1^{(n)}, \dots, \pi_d^{(n)}\}$ at time 0. For any positive s , let \mathcal{G}_s be the σ -algebra generated by the path of the random transposition Markov chain applied to these d permutations during time $[0, s]$. Let $\pi_i(s)$ denote the state of the i th permutation at time s . Then the vector-valued process $(\pi_1(s), \dots, \pi_d(s))$ is Markov with respect to this filtration.

Since T will be kept implicit in the analysis in this subsection, we will shorten $C_w(T, s)$ to $C_w(s)$. Fix an arbitrary $K \in \mathbb{N}$. We call a cycle (or a word) *short* if its length is at most K . Define the following $\{\mathcal{G}_s\}$ stopping times:

$$(31) \quad \tau_1 = \inf \{s \geq 0 : \text{a newborn short cycle shares a vertex with an existing short cycle}\},$$

$$\tau_2 = \inf \left\{ s \geq 0 : \sum_{w \in \mathcal{W}'_{2K}} C_w(s) > \sqrt{\log n} \right\},$$

$$\tau_3 = \inf \left\{ s \geq 0 : \sum_{w \in \mathcal{W}'_K} |C_w(s) - C_w(s-)| > 1 \right\}.$$

Note that τ_2 bounds cycles of length up to $2K$ and τ_3 rules out the possibility of the short cycle count jumping by more than one at any given moment in time.

Let $\bar{\sigma} = \min \{\tau_1, \tau_2, \tau_3\}$. Assume $\bar{\sigma} > 0$.

Definition 7. Let $\theta(s)$ denote the proportion of vertices counted in short cycles at time s . Thus, $\theta(0) \leq 2K\sqrt{\log n}/n$.

Step 1. Fix $s > 0$. Suppose there are N short cycles in the graph at time s . Let ξ_i be the stopping time when the i th short cycle vanishes. For $h \geq 0$, let $\mathcal{I}(s+h)$ be the vector of length N whose i th coordinate, $\mathcal{I}_i(s+h)$, is the indicator of the event $\{\xi_i \leq s+h\}$. Consider this now as a process in h that starts at the vector of all zeroes, and then, with progressing time, individual coordinates jump to one.

Lemma 16. *There exist a stopping time τ^* and a family of progressively measurable nonnegative processes $(\lambda_w^-(\cdot), w \in \mathcal{W}')$ such that the following happens.*

- (i) *Until $\tau^* \wedge \bar{\sigma}$, the process \mathcal{I} is a counting process, i.e., every coordinate increases exactly by one, and no two coordinates jump together.*
- (ii) *Suppose w is the word of the i th cycle. Then*

$$(32) \quad \mathcal{I}_i(s+h) - \int_0^{h \wedge \xi_i} \lambda_w^-(s+v) dv, \quad h \geq 0,$$

is a local martingale.

- (iii) *The death rates $\lambda_w^-(\cdot)$ satisfies the uniform estimate:*

$$(33) \quad \lambda_w^-(s) \in 2b(w) \left[1 - \frac{\sqrt{\log n}}{n}, 1 \right], \quad \text{for all } s \geq 0.$$

- (iv) *Let E^* denote an independent exponential random variable with mean $n/\log n$. Then, for any $t > 0$, $P(\tau^* \wedge \bar{\sigma} > t) \geq P(E^* \wedge \bar{\sigma} > t)$. That is, $\tau^* \wedge \bar{\sigma}$ stochastically dominates $E^* \wedge \bar{\sigma}$.*

Proof of Lemma 16. By the homogenous Markov property, it suffices to consider the case of $s = 0$. Thus consider all cycles existing at time 0. Suppose $\mathcal{I}(0) = x$. Consider the probability that the next transposition will turn the count to $x + e_i$, where e_i is the standard basis in \mathbb{R}^N and i is some coordinate which is currently 0. Consider the following collection of transpositions that can destroy the i th cycle while keeping others unchanged. Let w be the word of the i th cycle. As explained earlier in subsection 4.1 and Figure 2, a cycle with word w vanishes if one of the vertices (say u) incident at $b(w)$ many spots is involved in the transposition. Now given u , the proportion of v such that the transposition $[u, v]$ leads to $x + e_i$ and v is not a vertex of any short cycle is in between 1 and $1 - \theta(0)$ (recall $\theta(s)$ from Definition 7). Call this proportion $\delta_w(0)$. Let τ^* denote the first time that both vertices in the transposition are selected from the short cycles. Then, until $\tau^* \wedge \bar{\sigma}$, the coordinates of \mathcal{I} do not jump together and the infinitesimal death rate at time 0 makes sense:

$$(34) \quad \lambda_w^-(0) := \lim_{h \rightarrow 0} h^{-1} P[\mathcal{I}(h) = x + e_i, | \mathcal{G}_0, \mathcal{I}(0) = x] = 2b(w)\delta_w(0).$$

Similarly, the death rate for any other time $\lambda_w^-(s) := 2b(w)\delta_w(s)$ exists and satisfies

$$\left| \frac{\lambda_w^-(s)}{2b(w)} - 1 \right| \leq \theta(s) = O\left(\frac{\sqrt{\log n}}{n}\right), \quad \text{until } \tau^* \wedge \bar{\sigma}.$$

Moreover, by (34), each

$$\mathcal{I}_i(s+h) - \int_0^{h \wedge \xi_i} \lambda_w^-(s+v) dv$$

is a local martingale, where w is the word of the i th cycle.

Finally, we estimate the tails of τ^* . Both vertices are selected with a probability given by the square of the total number of vertices in short cycles over n . Thus, by our assumption on the short cycle count, we get

$$(35) \quad \lim_{h \rightarrow 0} \frac{1}{h} \mathbb{P}(\tau^* \leq h \mid \mathcal{G}_0) \leq \frac{\log n}{n}.$$

Let E^* denote an independent exponential random variable with mean $n/\log n$. By replacing \mathcal{G}_0 by any other \mathcal{G}_s , it follows from (35) that $\tau^* \wedge \bar{\sigma}$ stochastically dominates $E^* \wedge \bar{\sigma}$. This completes the proof of the lemma. \square

Step 2. We now consider the infinitesimal rates at which cycles are born.

Lemma 17. *Let $R_w(s)$ denote the number of cycles of word w that have ever existed during time $[0, s]$. Then, during $[0, \bar{\sigma})$, the vector-valued process $(R_w(\cdot), w \in \mathcal{W}'_K)$ is a counting process, i.e. each R_w jumps exactly by one, no two coordinates jump together. Moreover there exists nonnegative progressively measurable processes $(\lambda_w^+(s), w \in \mathcal{W}'_K)$ such that every $w \in \mathcal{W}'_K$*

$$R_w(s) - \int_0^s \lambda_w^+(u) du$$

is a local martingale. For $s \in [0, \bar{\sigma})$ we also have the following bound.

$$\frac{2b(w)}{h(w)} - C_1 \left(\frac{\sqrt{\log n}}{n} \right) \leq \lambda_w^+(s) \leq \frac{2b(w)}{h(w)}$$

where the positive constant C_1 depends only on d and K .

That the infinitesimal rates $\lambda_w^+(u)$ exist for all u is not hard to see. Consider the conditional law of the graph, conditioned on \mathcal{G}_0 . There is a certain number of transpositions which, if multiplied, increases any cycle count. Since, the vertices have independent exponential clocks, there exists an infinitesimal rate of increase of cycle counts $\lambda_w^+(u)$. We now compute it using the following lemma. Recall S_u (Definition 6) is the set of pre-cycles with word u .

Lemma 18. *Suppose we are given d permutations on n labels: π_1, \dots, π_d . Consider the graph generated by these permutations. Let A_j denote the set of vertices on this graph that lie on a cycle of length at most j . Given a short word w of length k , there exists a positive constant c_1 , depending on d and K ($1 \leq k \leq K$), such that*

$$2n \frac{b(w)}{h(w)} - c_1 A_k \leq \sum_{u \sim w} |S_u| \leq 2n \frac{b(w)}{h(w)},$$

where $u \sim w$ means that u and w are in the same equivalence class in \mathcal{W}'_k/D_{2k} .

Proof of Lemma 18. Fix a word w . Consider an $u \sim w$ and some $x \in [n]$. For $|S_u| > 0$, we need u_1 and u_k to have the same sign. Therefore, the number of such u 's is $2b(w)/h(w)$.

For such an u we claim that there is at most one pre-cycle with word u whose j th vertex is x for $j = 0, 1, \dots, k$. This is true, since the trail of the only possible pre-cycle is given by

$$\begin{aligned} s_j &= x, s_{j+1} = u_{j+1}(s_j), s_{j+2} = u_{j+2}(s_{j+1}), \dots, \\ s_{j-1} &= u_j^{-1}(s_j), s_{j-2} = u_{j-1}^{-1}(s_{j-1}), \dots \end{aligned}$$

Such a trail need not be valid due to repeated vertices (including the first and the last which will make it closed). Hence, the qualifier ‘at most’. However, suppose that there are two indices $i_1 < i_2$ such that $s_{i_1} = s_{i_2}$, then the closed trail from i_1 to i_2 forms a cycle of size $i_2 - i_1$ which is at a graph distance at most k from x .

There are at most $k(2d)^k |A_k|$ many vertices in the graph that are at a distance at most k from a cycle of length at most k . For all other vertices, the trail given above is not closed, has no repeated vertices and is a pre-cycle with word u .

Hence, if we define $c_0 = K(2d)^K$, then, for a fixed $u \sim w$ and any $1 \leq k \leq K$, we have

$$(k+1)(n - c_0 |A_K|) \leq \sum_{x \in [n]} \sum_{j=0}^k \sum_{s \in S_u} 1_{\{s_j=x\}} \leq (k+1)n.$$

On the other hand, interchanging the order of summation above we get

$$\sum_{x \in [n]} \sum_{j=0}^k \sum_{s \in S_u} 1_{\{s_j=x\}} = \sum_{s \in S_u} \sum_{j=0}^k \sum_{x \in [n]} 1_{\{s_j=x\}} = (k+1) |S_u|.$$

Combining the two estimates we get

$$(n - c_0 |A_K|) \leq |S_u| \leq n.$$

Summing up over all possible $u \sim w$ proves the lemma with $c_1 = 2Kc_0$. \square

Proof of Lemma 17. We now return to computing the infinitesimal rate of increase of cycles. As before, it suffices to calculate the rate at time 0 and appeal to the Markov property to argue for every other time. At time 0, every possible transposition $[x, y]$, where (x, y) is ordered, occurs with rate $1/n$. Fix a short word w . We ask: how many transpositions will turn a pre-cycle to a cycle with word w ?

To answer this question suppose the transposition is $[x, y]$ for $x \neq y$. Let $u \sim w$ be a word with the first and last letters having the same sign. Then, we are limited to all pre-cycles with word u that either start with x and end with y , or start with y and end with x . Out of these, every pair of equivalent pre-cycles produce the same cycle by the same transposition. Thus, every such pair is to be counted once. Hence, the rate at which cycles with word w get created is given by

$$\frac{1}{2n} \sum_{(x,y) \in [n]^2, x \neq y} \sum_{u \sim w} \sum_{s \in S_u} [1_{\{s_0=x, s_k=y\}} + 1_{\{s_0=y, s_k=x\}}]$$

As before, we interchange the order the summation we get

$$\begin{aligned} & \frac{1}{2n} \sum_{u \sim w} \sum_{s \in S_u} \sum_{x \in [n]} \sum_{y \neq x} [1_{\{s_0=x, s_k=y\}} + 1_{\{s_0=y, s_k=x\}}] \\ (36) \quad &= \frac{1}{2n} \sum_{u \sim w} \sum_{s \in S_u} \sum_{x \in [n]} [1_{\{s_0=x\}} + 1_{\{s_k=x\}}] = \frac{1}{n} \sum_{u \sim w} |S_u|. \end{aligned}$$

The final number has been counted in Lemma 18; we get

$$(37) \quad \lim_{h \rightarrow 0+} h^{-1} \mathbb{E}[R_w(s+h) - R_w(s) \mid \mathcal{G}_s] = \lambda_w^+(s) = \frac{1}{n} \sum_{u \sim w} |S_u|.$$

Note that the bound above works for all time points, not just those before $\bar{\sigma}$. The proof of the lemma now follows from Lemma 18 and the fact that upto $\bar{\sigma}$, $\sqrt{\log n}$ is an upper bound of the number of short cycles. Also the process is a counting process until $\bar{\sigma}$ by definition. \square

Remark 3. However note that the rate at which more than one cycle gets created is the rate at which a transposition $[x, y]$ occurs where the pair x, y belong to two precycles which mean that both x, y lie on a cycle of size at most $2K$. This is the first time we need to use a bound on cycles of length $2K$ to bound rates concerning cycles of length at most K . This explains the definition of τ_2 . Now by definition of τ_2 upto $\bar{\sigma}$ there are at most $O(\log n)$ (the constant in the O term depends only on K), such transpositions. Also such a transposition can create at most K many new cycles since every point is on at most K precycles as discussed in the proof of Lemma 18. Thus denoting for all $s \in [0, \bar{\sigma})$ the rate at which exactly one cycle gets created by $\tilde{\lambda}_w^+(s)$ we have

$$\tilde{\lambda}_w^+(s) = \lambda_w^+(s) - O\left(\frac{\log n}{n}\right)$$

where $\lambda_w^+(s)$ appears in (37) and the constant in the O term depends only on K .

Step 3. We define new processes by extending the birth rates $\tilde{\lambda}_w^+$ and death rates λ_w^- (Lemma 16) from $[0, \bar{\sigma})$ to all times in $[0, \infty)$ by defining

$$\tilde{\lambda}_w^+(s) = \frac{2b(w)}{h(w)}, \quad \lambda_w^-(s) = 2b(w), \quad s \geq \bar{\sigma}, \quad w \in \mathcal{W}'_K.$$

Consider the following time changes that are measurable with respect to the predictable σ -algebra:

$$\Gamma_w^+(s) = \int_0^s \tilde{\lambda}_w^+(v) dv, \quad w \in \mathcal{W}'_K, \quad \Gamma_w^-(s) = \int_0^s \lambda_w^-(v) dv.$$

We now extend the counting process R from $[0, \bar{\sigma})$ to the entire time axis. Let us for the moment call it \tilde{R} . During time $[0, \bar{\sigma})$ the coordinates of \tilde{R} are exactly the same as R as in Lemma 17. At $\bar{\sigma}$ in the actual process, R_w can jump by more than 1, however \tilde{R}_w in that case does not jump at all. Lastly during $[\bar{\sigma}, \infty)$, each \tilde{R}_w continues as an independent Poisson process with rate $2b(w)/h(w)$. Clearly, this extended process is a counting process on $[0, \infty)$. Thus for all $w \in \mathcal{W}'_K$ and $s \in [0, \bar{\sigma}]$

$$\tilde{R}_w(s) = \sum_{t \leq s} \Delta R_w(t) \mathbf{1}(\Delta R_w(t) = 1)$$

where for any t

$$\Delta R_w(t) = R_w(t) - R_w(t_-)$$

is the jump size at t . The above sum makes sense since the process R_w has only finitely many jump points in any finite time interval. Let $A_w^+(\cdot)$ be the compensator for the coordinate process $\tilde{R}_w(\cdot)$. Then by construction $A_w^+(s) = \Gamma_w^+(s)$ for $s \in [0, \bar{\sigma})$. To see this notice that by definition the rate of increase of the process \tilde{R} at any time is the rate at which one cycle is produced in the process R . This is $\tilde{\lambda}_w^+(s)$ for all times $s < \bar{\sigma}$ and equal to $2b(w)/h(w)$ on the interval $[\bar{\sigma}, \infty)$.

Lemma 19. *It is possible to extend our current probability space to define a Poisson point processes Q on $\mathcal{W}'_K \times [0, \infty)$ such that the following holds. The intensity measure for Q on $\mathcal{W}'_K \times (0, \infty)$ is the \mathcal{W}'_K -fold product of Leb. At $\mathcal{W}'_K \times \{0\}$ the count coincides with the number of short cycles at time 0. Moreover, if we consider*

the extended counting process $(\tilde{R}_w(\cdot), w \in \mathcal{W}'_K)$ in Lemma 17 as a point process on $\mathcal{W}'_K \times [0, \infty)$ as described above, then

$$\tilde{R}_w(s) = Q_w(A_w^+(s)), \quad s \in [0, \infty), \quad w \in \mathcal{W}'_K.$$

Proof of Lemma 19. This is a consequence of the Watanabe-Meyer theorem [Wat64, Mey71] that every counting process is a time-changed Poisson process. For a proof see [BN88]. The compensator has already been defined above. The extension of the space is necessary to extend the point process on $[\bar{\sigma}, \infty)$. \square

Step 4. We can now extend the point process $(R_w(\cdot))$ to an extension of the point process $\chi_K(T)$ in Proposition 15 by noting the lifetime of each cycle. Formally, for every atom (w, s) in R , extend it to (s, v, w) , where v is the length of time that the short cycle which gets created at time s exists. If a cycle with word w exists beyond $\bar{\sigma}$ or gets born after $\bar{\sigma}$, the (possibly excess) lifetime will be i.i.d. $\text{Exp}(2b(w))$. We will refer to this extended process by $\bar{R}(s, v, w)$. Thus $\chi_K \equiv \bar{R}$ during $s \in [0, \bar{\sigma})$.

Lemma 20. *As T goes to infinity, \bar{R} converges to a PPP on $[0, \infty) \times [0, \infty) \times \mathcal{W}'_K$ that has independent $\text{Poi}(1/h(w))$ many atoms for w at time $s = 0$, births at rate $2b(w)/h(w) \times \text{Leb}$ and a lifetime of $\text{Exp}(2b(w))$.*

Proof of Lemma 20. The proof follows from the coupling in Lemma 19, the estimate in Lemma 16, and Lemma 21 below. Since the arguments are standard we outline the major steps and skip the details.

It suffices to argue that if we take finitely many disjoint intervals on the time line, then there are independent Poisson many births of the correct rate, and that, each such newborn cycle survives an independent exponential amount of time. As $T \rightarrow \infty$, $M_T = n$ goes to infinity in probability. Therefore, by Lemma 21 below, uniformly over compact sets in s ,

$$\lim_{T \rightarrow \infty} \Gamma_w^+(s) \stackrel{p}{=} \frac{2b(w)s}{h(w)}, \quad \forall w \in \mathcal{W}'.$$

The coupling in Lemma 19 then gives us weak convergence on the birth counts. Conditioned on the birth counts, Lemma 16 allows us to couple each lifetime with independent exponentials. The lemma now follows from the explicit error bounds given in (33). \square

Step 5. So far we have coupled χ_K with another point process \bar{R} whose weak limit is the limiting field. The two fields are identical during time interval $[0, \bar{\sigma})$. Hence, Proposition 15 follows once we argue the following.

Lemma 21.

$$(38) \quad \lim_{n \rightarrow \infty} P(\bar{\sigma} \leq S) = 0, \quad \text{for all } S > 0.$$

Proof of Lemma 21. Fix $\epsilon > 0$. Consider the stopping times defined in (31). First consider the event $\{\bar{\sigma} > 0\}$. The event only depends on the graph $G(n, 2d)$. We know from [JP14, Corollary 25] that, as n goes to infinity, asymptotically almost surely no two cycles in the graph share a vertex. We also know that the asymptotic law of the vector $(C_w(0), w \in \mathcal{W}'_{2K})$ is the product of independent $\text{Poi}(1/h(w))$

with a convergence in total variation. In fact, we know from [JP14, Corollary 24] that

$$P\left(\sum_{w \in \mathcal{W}'_{2K}} C_w(0) > \frac{1}{2}\sqrt{\log n}\right) - P\left(\gamma_{2K} > \frac{1}{2}\sqrt{\log n}\right) \leq c \frac{(2d-1)^{2K}}{n},$$

where c is an absolute constant and γ_{2K} is a Poisson random variable with mean

$$\sum_{w \in \mathcal{W}'_{2K}} \frac{1}{h(w)} = \sum_{j=1}^{2K} \frac{a(d, j)}{2j} \leq \sum_{j=1}^{2K} a(d, j) \leq C_0(2d-1)^{2K},$$

where C_0 is a constant that depends only on d . Therefore, by taking a large enough n we can guarantee that the event

$$(39) \quad A_0 := \left\{ \sum_{w \in \mathcal{W}'_{2K}} C_w(0) < \frac{1}{2}\sqrt{\log n} \right\} \cap \{\tau_1 > 0\}$$

occurs with probability at least $1 - \epsilon/2$. Since $\tau_3 > 0$, the event $\{\bar{\sigma} > 0\}$ has probability at least $1 - \epsilon/2$.

Start from an initial configuration that satisfies A_0 . The stopping time τ_3 guarantees that the process of short cycle counts do not jump by more than one. Consider the possible cases when τ_3 happens by defining three other stopping times.

- (i) τ_4 is the first time two short cycles with words w, w' (possibly same) appear simultaneously.
- (ii) τ_5 is the first time two short cycles with those words disappear simultaneously.
- (iii) τ_6 is the first time when one short cycle appears while another disappears.

Consider $\tilde{\tau} = \tau_1 \wedge \tau_2 \wedge \tau_4$. We first evaluate the probability that $\{\tilde{\tau} > S\}$ for any $S > 0$. Call a transposition at time s *bad* if multiplication by that transposition will lead to $\tilde{\tau}$. Hence any time $s \in [0, \tilde{\tau})$ it suffices to count the number of bad transpositions. This is a computation similar to the proof of Lemma 17 and 18.

Let A_1 be the event that there are less than $\sqrt{\log n}/2$ many new births of cycles of size at most $2K$ during $[0, S]$. Thus, under $A_0 \cap A_1$, the cycle counts of those cycles never exceed $\sqrt{\log n}$. We estimate the probability of A_1 . At any given s , we bound the total number of pre-cycles that can possibly give us one or more cycles (simultaneously) of size at most $2K$.

Given any word, it follows from (37) and Lemma 18 that the expected total number of cycles of word w created during time $[0, S]$ is at most $2Sb(w)/h(w)$. For a word of length at most $2K$, we can bound this mean by $2KS$. There are at most $(2d)^{2K+1}$ many words of length at most $2K$. Thus the expected number of cycles with such words born during $[0, S]$ is bounded above by $2KS(2d)^{2K+1}$. Hence, by Markov's inequality, one can choose n large enough such that the probability of more than $\sqrt{\log n}/2$ births is at most $\epsilon/4$. By a union bound,

$$(40) \quad P(A_0 \cap A_1) \geq 1 - 3\epsilon/4.$$

Assume that $A_0 \cap A_1$ holds. Thus $\tau_2 > S$. Consider τ_4 . Denote the state of the permutations at (τ_4-, τ_4) by $(\pi_i-, i \in [d])$ and $(\pi_i, i \in [d])$. Here τ_4- refers to left limit of the chain at τ_4 . Now observe the change in reversed time. Let σ be the transposition that occurs at τ_4 . By taking inverses, each π_i- can be

obtained from π_i by left multiplication by the same transposition σ . New cycles are formed by multiplying pre-cycles with σ . Therefore we see that two cycles can appear simultaneously at τ_4 if and only if at τ_4 the vertices formed a cycle of size $|w| + |w'|$ and the transposition chooses two elements from its vertices. Since this cycle can be of size at most $2K$ and we have assumed $\tau_2 > S$, the rate at which such a transposition occurs is bounded by

$$(41) \quad \frac{1}{n} \left(\sum_{w \in \mathcal{W}'_{2K}} C_w \right)^2 \leq O\left(\frac{\log n}{n}\right).$$

where the constant in the order term only depends on K .

Now consider the rate which new short cycles share vertices with existing short cycles. This happens if there is a short pre-cycle that shares a vertex with an existing cycle. The number of existing short cycles is bounded by $\sqrt{\log n}$. Each has at most K vertices. It has been shown in Lemma 18 that each vertex can lead at most K^2 pre-cycles with a given short word. Hence, the rate is bounded above by $c_1 \log n/n$, where c_1 is a constant depending on d and K .

Hence, during $[0, S]$, the random variable $\tau_1 \wedge \tau_4$ stochastically dominates an exponential random variable of rate $c_2 \log n/n$, where c_2 is some constant depending only on d and K . Thus one can take n large enough to guarantee that

$$\mathbb{P}(\tilde{\tau} > S) \geq 1 - \frac{7\epsilon}{8}.$$

Now consider τ_5 . Until $\tilde{\tau}$ no two short cycles share a vertex. Hence τ_5 happens only if the transposition selects both vertices counted in short cycles and the cycles merge. The probability of such a transposition is bounded by $\log n/n$. Thus, comparing with an exponential with rate $\log n/n$, we see that one can take n large enough to guarantee that

$$\mathbb{P}(\tilde{\tau} \wedge \tau_5 > S) \geq 1 - \frac{9\epsilon}{10}.$$

The case of τ_6 is similar. The transposition has to involve a vertex of a small cycle and another of a small pre-cycle. In any case, the rate of such transpositions is again of the order of $\log n/n$. Thus, one can take n large enough to guarantee

$$\mathbb{P}(\bar{\sigma} > S) = \mathbb{P}(\tilde{\tau} \wedge \tau_5 \wedge \tau_6 > S) \geq 1 - \epsilon.$$

Since ϵ is arbitrary, this proves our claim. \square

The next subsection proves Proposition 14 using Proposition 15 by proving that the projection of the backward chinese restaurant process on the space of cycles converge to the halving chain defined in Definition 2. A one dimensional version of this result is proved in the proof of Theorem 16 in [JP14] and some of the basic arguments appearing there are used in the following proof.

4.4. Convergence to halving chains. Fix time $s > 0$. Consider the collection of cycles present at any time during $[0, s]$ at dimension T . We consider their evolution backward in dimension. We claim that these evolve as independent halving chains. Notice that we do not mention their lifetime. This is because, by construction, if the same cycle exists at two time points, their backward evolution is identical. Let us outline the argument Consider the graph valued process backward in dimension

$$\overleftarrow{G}_v(u) = G(T - u, v), \quad v \in [0, s], \quad u \geq 0.$$

Fix $K \in \mathbb{N}$. Choose a large positive integer $L \gg K$ and ignore all of $\overleftarrow{G}_v(0) = G(T, v)$, $v \in [0, s]$, except for the subgraph consisting of cycles of size L and smaller. Call this graph $\overleftarrow{\Gamma}_v(0)$. Consider now the evolution, backward in dimension, of the graphs $\overleftarrow{\Gamma}_v(\cdot)$, $0 \leq v \leq s$.

Define an event \mathcal{E} which states that no two of the finitely many cycles of size at most L that are created in dimension T and during time $[0, s]$ share a vertex. Then, our proof goes by shown (i) \mathcal{E} holds asymptotically almost surely, and (ii) under \mathcal{E} the analysis of the backward processes $\overleftarrow{\Gamma}_v(\cdot)$ is trivial. This suffices then by a Lemma 7 type argument which says large cycles do not quickly shrink to small cycles in the halving chain.

Claim (i) is almost Lemma 21. The only difference is that it is possible for two cycles existing at disjoint intervals of time to have a common vertex. But, an easy extension to the same argument covers this case.

Let us now explain (ii). We ignore the vertex labels and consider every vertex in the graph to have an exponential one clocks attached to it. Backward in dimension, whenever the clock of a vertex rings, we remove that vertex from every permutation. By our construction, the same vertex (or, more precisely, its image) is removed simultaneously from every point in time. The remarkable fact is that, under \mathcal{E} , each cycle evolves independently as a halving chain for every T .

Let R be the rectangle $[-T_0, 0] \times [0, S_0] \times \mathcal{W}'_K$. Now,

$$\begin{aligned} &P\left(C_w(T+t, s) = N_w^{(T, L)}(t, s), \forall (t, s, w) \in R\right) \\ &\geq P\left(C_w(T+t, s) = N_w^{(T, L)}(t, s), \forall (t, s, w) \in R \mid \mathcal{E}\right) P(\mathcal{E}). \end{aligned}$$

Hence we will be done once we show that, for any $\delta > 0$ there exists large enough L such that,

$$(42) \quad \limsup_{T \rightarrow \infty} P\left(C_w(T+t, s) = N_w^{(T, L)}(t, s), \forall (t, s, w) \in R \mid \mathcal{E}\right) > 1 - \delta.$$

The argument is similar to the proof of Lemma 7 and hence we use similar notation. Let $E_T(L)$ be the event that some cycle of length $l > L$ that exists at dimension T and anywhere in time $[0, S_0]$ at least $l - K$ of its shrinking vertices are deleted by dimension $T - T_0$.

For $l > L$, word $w \in \mathcal{W}_l/D_{2l}$, and $I \subseteq [l]$, such that $|I| = l - K$, $w_i = w_{i+1}$ for $i \in I$, let $F(w, I)$ denote the event that a cycle of word w that exists at some point in $\{T\} \times [0, S_0]$ shrinks all the vertices in I by dimension $T - T_0$. Thus exactly as (14)

$$(43) \quad P(E_T(L)) \leq \sum_{w, I} P[F(w, I)].$$

And then similarly as in (15) we have

$$(44) \quad P[F(w, I)] \leq (1 - e^{-T_0})^{l-K} E(G_w).$$

where G_w denotes the number of cycles of word w that ever exists during $\{T\} \times [0, S_0]$.

The following lemma is needed to finish the current argument.

Lemma 22. *For any w and any $T > 0$, we must have $E(G_w) \leq 2S_0 |w| + 1$.*

Keeping the proof of this lemma for last, let us continue with the previous argument. Using Lemma 22 as in (18) we have

$$(45) \quad P[E_T(L)] \leq (2d)^K \sum_{l=L+1}^{\infty} (2S_0 l + 1) l^K (1 - e^{-T_0})^{l-K}.$$

The right side of the above bound is summable. Hence, one can find L large enough such that it is smaller than any $\delta > 0$ proving (42).

Proof of Lemma 22. Given a word w the number of cycles to ever exist during $\{T\} \times [0, S_0]$ consists of two kinds: (i) those that exist at $(T, 0)$ and (ii) those that are created during $\{T\} \times (0, S_0]$. The expected value of the first kind is at most one as argued in the second displayed equation on [JP14, p. 19]. So, we focus on the second kind.

Consider the counting in (36) for a word w that need not be short. In any case, (37) continues to hold:

$$\lim_{h \rightarrow 0} h^{-1} \mathbb{E}[R_w(s+h) - R_w(s) \mid \mathcal{G}_s, R_w(s) = x] \leq \frac{2b(w)}{h(w)}.$$

Therefore, for any w , during time $(0, S_0]$ the expected total number of w cycles created during that time is dominated by

$$\frac{2b(w)}{h(w)} S_0 \leq 2|w| S_0.$$

This completes the argument. \square

4.5. Weak convergence of cycle counts.

Finally we prove Theorem 1.

Proof of Theorem 1. Consider the set-up of Proposition 14, in particular, the rectangle $R = [-T_0, 0] \times [0, S_0] \times \mathcal{W}'_K$. We have four different random fields of interest.

- (i) The field of cycles: $(C_w(T+t, s), (t, s, w) \in R)$.
- (ii) The field $(N_w^{(T,L)}(t, s), (t, s, w) \in R)$ with non-limiting initial conditions $\chi_L(T)$.
- (iii) We now define the field $(N_w^L(t, s), (t, s, w) \in R)$ of limiting cycle counts with the limiting initial condition χ_L .
- (iv) And the actual limiting field $(N_w(t, s), (t, s, w) \in R)$ defined in Definition 3.

Proposition 14 and Slutsky's theorem implies that any weak limit of fields (i) and (ii) must be the same. We also know from Proposition 10 that the field (iii) converges almost surely to (iv) as L tends to infinity. Hence, it suffices to show that, as T tends to infinity, the field (ii) converges weakly to (iii).

We know that $\chi_L(T)$ converges weakly to χ_L as random Radon measures on the space $[0, \infty) \times (0, \infty) \times \mathcal{W}'_L$. By Skorokhod's theorem, one can construct a probability space and copies of $\chi_L(T)$ and χ_L such that this convergence holds almost surely. We will restrict ourselves to this constructed space and construct couplings.

On this space one can find T' large enough such that, for all $T \geq T'$, the total mass of $\chi_L(T)$ on every $w \in \mathcal{W}'_L$ is the same as that of χ_L . This is because, almost surely, there are only finitely many of them. Assume, henceforth, $T \geq T'$.

Informally, the same number of cycles of each word get created under $\chi_L(T)$ and χ_L . For every word w enumerate these atoms according to the times of their births. Then, take an identical set of independent halving chains starting from w

and attach to the two sets of atoms in the same order. Thus, we have a collection of independent Markov chains whose birth times and lifetimes are *slightly off*, however, the correct alignment renders their paths identical. This produces a coupling of field (ii) and (iii).

However, as T goes to infinity, the finite vector of birth times and lifetimes of every cycle in $\chi_L(T)$ is a continuous function of the point process, converges almost surely to that of χ_L . By construction of our coupling this causes the fields (ii) to converge to (iii) pointwise in the rectangle R . This implies converges in the space D_2 trivially (see [Neu71, eqn. (2.3), (2.4)]).

The claimed weak convergence now follows. \square

5. LINEAR EIGENVALUE STATISTICS

Let us recall some of the basic facts established in [DJPP12, Section 3, 5] and [JP14, Section 5] that connect linear eigenvalue statistics with cycle counts. Recall G_n is a random regular graph of degree $2d$ on n vertices. A closed non-backtracking walk is a walk that begins and ends at the same vertex, and that never follows an edge and immediately follows that same edge backwards. If the last step of a closed non-backtracking walk is anything other than the reverse of the first step, we say that the walk is *cyclically non-backtracking* (CNBW). Cyclically non-backtracking walks on G_n are exactly the closed non-backtracking walks whose words are cyclically reduced. Let $\text{CNBW}_k^{(n)}$ denote the number of closed cyclically non-backtracking walks of length k on G_n .

Let $\{T_n(x)\}_{n \in \mathbb{N}}$ be the Chebyshev polynomials of the first kind on the interval $[-1, 1]$. We define a set of polynomials

$$\begin{aligned}\Gamma_0(x) &= 1, \\ \Gamma_{2k}(x) &= 2T_{2k}(x) + \frac{2d-2}{(2d-1)^k}, \quad \forall k \geq 1, \\ \Gamma_{2k+1}(x) &= 2T_{2k+1}(x), \quad \forall k \geq 0.\end{aligned}$$

Let A_n be the adjacency matrix of G_n , and let $\lambda_1 \geq \dots \geq \lambda_n$ be the eigenvalues of $(2d-1)^{-1/2}A_n/2$. Then

$$(46) \quad \sum_{i=1}^n \Gamma_k(\lambda_i) = (2d-1)^{-k/2} \text{CNBW}_k^{(n)}.$$

Now, for any cycle in G_n of length $j|k$, we obtain $2j$ non-backtracking walks of length k by choosing a starting point and direction and then walking around the cycle repeatedly. It follows from [DJPP12, Corollary 18], that for fixed d and r , all cyclically non-backtracking walks of length r or less have this form with high probability. Thus the random vectors $(\text{CNBW}_k^{(n)}, 1 \leq k \leq r)$ and $(\sum_{j|k} 2jC_j^{(n)}, 1 \leq k \leq r)$ have the same limiting distribution, and the problem of finding the limiting distributions of polynomial linear eigenvalue statistics is reduced to finding limiting distributions of cycle counts.

Most of this section is devoted to showing that a similar statement holds for the entire two parameter field $G(\cdot, \cdot)$. Let $\text{CNBW}_k(\cdot, \cdot)$ denote the corresponding field of cyclically non-backtracking walks of length k . As in [JP14] call a CNBW *bad* if

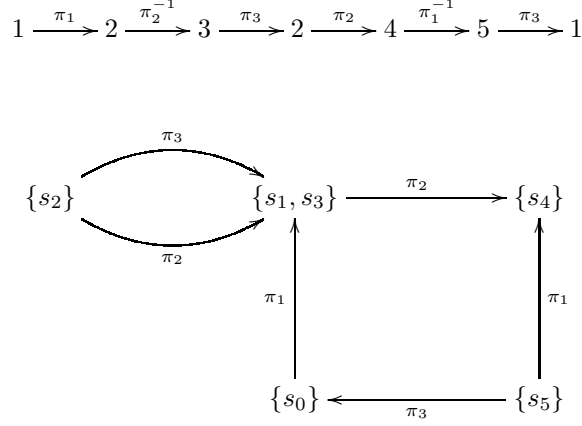


FIGURE 4. A bad CNBW trail and its category graph

it is anything other than a repeated walk around a cycle. Formally we prove the following proposition.

Proposition 23. *Fix any $T_0, S_0, \epsilon > 0$ and $K \in \mathbb{N}$. Consider T for which the rectangle $R_0 := [T - T_0, T] \times [0, S_0]$ is a subset of $[0, \infty)^2$. Then, for all large enough T , the probability that there is any bad CNBW of length at most K in the field of graphs $G(t, s)$, $(t, s) \in R_0$, is at most ϵ .*

To prove the above proposition we recall the definition of categories of trails from [DJPP12, Section 3]. Recall the definition of trails from Definition 6. In this section we will use an expanded definition that allows vertices to repeat. If we start at a vertex of a CNBW and walk around it, we get a closed trail with possible repeated vertices. Given such a trail its category is a directed, edge-labeled graph depicting the overlap of trails. Essentially one gets a category from a trail by gluing the vertices with the same label. See Figure 1 in [DJPP12] or Figure 7 in [LP10] for the general definition of categories of a list of trails.

This new graph is called the category of the trail. See Figure 4 for an example. One can now inquire how many trails of a certain category appears in a given graph. The crucial property of categories is the difference between the number of edges and vertices. We introduce the new definition.

Definition 8. For a category Γ its characteristic will refer to the quantity $e - v$, where e is the number of edges of Γ and v is the number of vertices.

The characteristic of a category determines how likely it is to appear in $G(n, 2d)$. We will refer to the following bound from [DJPP12, Lemma 14]. Let $X_\Gamma^{(n)}$ be the number of trails of category Γ that appear in $G(n, 2d)$ for $n > e$. Then

$$(47) \quad \mathbb{E} \left(X_\Gamma^{(n)} \right) \leq \frac{1}{[n - k]_\gamma},$$

where γ is the characteristic of Γ and $[x]_j$ is the falling factorial $[x]_j = x(x - 1) \cdots (x - j + 1)$.

Lemma 24. *For any $S_0 > 0$ and any $\epsilon > 0$, for all large enough T , the probability that there exists a bad CNBW of length at most K in some $G(T, s)$, $0 \leq s \leq S_0$, whose category has characteristic 2 or higher is at most ϵ .*

Proof. For a given graph G with law $G(n, 2d)$, the probability that it has a bad CNBW of characteristic 2 or higher is $O(n^{-2})$. This follows from [DJPP12, Proposition 15] when we start the summation in the statement from $i = 2$ (instead of $i = 1$). The constant in the big- O depends only on d and K .

Suppose at dimension T the graphs have n vertices. Given S_0 , there are a Poisson with mean nS_0 many transpositions that occur during time $[0, S_0]$. Thus with probability at least $1 - \frac{\epsilon}{2}$, $O(n)$ many distinct graphs that ever exist during time $[0, S_0]$. By a union bound, the probability that any of them will have a bad CNBW of characteristic 2 or higher is $O(1/n)$. We skip the easy details. \square

Definition 9. For any pair $(l, j) \in \mathbb{N}^2$ we say that a graph is (l, j) tangle free if any two cycles of length at most l are at a graph distance at least j from one another.

Remark 4. The concept of *tangle-free-ness* has been given other similar but slightly different definitions. See, for example, [Fri08]. Our definition is relevant only to the following argument.

Lemma 25. *If a graph is (l, l) tangle free then it has no bad CNBW of length l or less.*

Proof. Consider a trail corresponding to a bad CNBW of length l (or less) by starting from a vertex and walking along the walk. The lemma is clear if one imagines the category graph for this trail.

This graph has to have a cycle, since otherwise it is not a cyclically non backtracking walk. However, the graph cannot be the cycle itself, since, otherwise the CNBW will be a repeated cycle which cannot be bad. Thus, consider the first cycle one encounters following the trail. That is, the first time we encounter a vertex encountered before. If one *erases* this cycle from the graph, the remaining is non-empty. By erasing a cycle we mean the following procedure: consider the part of the trail that constitutes the cycle. Remove all vertices and edges, except the first vertex where the cycle starts and the edge that connects the final vertex (which is the same as the first) to the next vertex (which is not in the cycle).

Now, we claim that the remaining graph must also have a cycle. This is evident because erasing the cycle did not destroy the property that the remaining walk is closed. The two cycles thus found have length at most l and a graph distance at most $l - 2$. If it exists in the graph, it contradicts the (l, l) tangle free property. \square

The plan of the proof of Proposition 23 is the following. Consider a bad CNBW of length at most K that exists at some point in the rectangle R_0 . Consider its trail and the corresponding category graph. Fix the time axis, and move forward in dimension. As a new vertex gets added to the graph, it is possible that it will have edges with the existing vertices. Suppose the new vertex is k . According to the rule of CRT, an existing edge

$$i \xrightarrow{\pi_i} j \quad \text{or} \quad i \xrightarrow{\pi_i^{-1}} j$$

can only change into

$$(48) \quad i \xrightarrow{\pi_i} k \xrightarrow{\pi_i} j \quad \text{or} \quad i \xrightarrow{\pi_i^{-1}} k \xrightarrow{\pi_i^{-1}} j, \quad \text{respectively.}$$

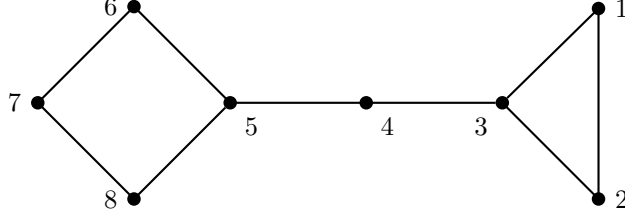


FIGURE 5. The trail $1 \rightarrow 2 \rightarrow 3 \rightarrow 4 \rightarrow 5 \rightarrow 6 \rightarrow 7 \rightarrow 8 \rightarrow 5 \rightarrow 4 \rightarrow 3 \rightarrow 2 \rightarrow 1$ is a double lollipop. The edge labels are not shown.

Introduce the new vertex into the walk by including it as above by choosing one of the possible edges it belongs to.

Modify the category graph by adding the new vertex and the new edge. The important observation is that the characteristic of the category of the trail is non-decreasing. This is because when a new vertex gets added to the existing trail a new edge also gets added, and therefore the trail remains a bad CNBW.

Thus, for a bad CNBW ω of characteristic γ , there must exist another bad CNBW ω' of characteristic at least γ at dimension T such that the former can be obtained from the latter by deletion of vertices. This is a very similar situation to the arguments in Subsection 4.4 where we looked into larger cycles at dimension T shrinking to smaller cycles at a lower dimension. In short, one can expect an $O(1)$ many vertices to be included in the bad CNBW as the dimension increases to T . Suppose the length of ω is j and that of ω' is l . There are $i = l - j$ many new vertices added to the trail. The proof of Lemma 25 shows that there must be at least two cycles in ω that are at distance at most $j - 2$ from each other.

Some of the i many extra vertices will possibly increase the size of these cycles, some will possibly increase the graph distance. In any case, at dimension T we will have two cycles of length at most l that are at a graph distance at most $l - 2$. Thus, the (l, l) tangle free condition will be violated at some time point at dimension T .

Our proof has two parts. One, to show that for any fixed l , with high probability, the (l, l) tangle free condition holds at dimension T during time $[0, S_0]$. Two, to show that for any $k \in \mathbb{N}$, there is a large $l \in \mathbb{N}$, such that with high probability all bad CNBWs at a lower dimension of length at most k have shrunk from bad CNBWs of length at most l at dimension T , for all large enough T .

Lemma 26. *Fix any $\epsilon > 0$ and any $L \in \mathbb{N}$. Then, for all large enough T , we have*

$$\mathbb{P}(G(T, s) \text{ is } (L, L) \text{ tangle free for all } s \in [0, S_0]) \geq 1 - \epsilon.$$

The above statement is essentially the same as modifying the definition of the stopping time τ_1 in (31) to be the time till which all cycles of length at most L have disjoint L -neighborhoods. The proof hence is essentially the same as the proof of Lemma 21. We include the proof for clarity and to introduce some new structures, needed for the proof of the next lemma.

Proof. Consider any three integers i, j, k , each less than L . Consider a category graph that connects two cycles of sizes i and j by a single path of length $k \geq 0$. We call this graph (or the trail) a *double-lollipop* for obvious reason. See Figure 5

for a case of $i = 4, j = 3, k = 1$. Given a double-lollipop associate a word with it by choosing any hamiltonian path and traversing the edges along it. The importance of the double-lollipops come from the Lemmas 24 and 25. A graph is not tangle-free if and only if either (i) two short cycles intersect at more than one vertex, or (ii) there are two short cycles joined by a short path (i.e., a double-lollipop). Case (i) is unlikely by Lemma 24. It remains is to argue that case (ii) is also unlikely.

It is known from [DJPP12, Section 5] that, for T large enough, the probability that $G(T, 0)$ is not (L, L) tangle free is at most $\epsilon/2$. So, we restrict attention to new double-lollipops that are ever born during $(0, S_0]$. The argument is similar to that of pre-cycle counts in Lemma 18, Lemma 17, and the proof of Lemma 21. The following is the idea.

Suppose we are given a trail corresponding to a double-lollipop with word w . Then the length of w is $i + j + k \leq 3L$. We define a pre-lollipop to be a trail (or a collection of two trails) that is obtained by pre-multiplying a transposition with the trail of the double-lollipop. Suppose we show, with high probability, that there is $O(\log n)$ many pre-lollipops that ever exist during time $[0, S_0]$, independent of n . Then, only $O(\log n)$ many vertices are included in these pre-lollipops. Thus, if a double-lollipop ever has to occur during time $[0, S_0]$, the required transposition must involve two vertices from these pre-lollipops. But, this is highly unlikely, since each transposition occurs with probability $2/n^2$ and only a Poisson with mean nS_0 ever transpositions occur during $[0, S_0]$.

Consider the stopping time $\bar{\sigma}$ similar to Lemma 21, except in (31) we consider short cycles to be those with size at most L . We start by showing that the number of pre-lollipops born during $[0, \bar{\sigma})$ is of order $\log n$. Fix a double-lollipop and imagine the different pre-lollipops that are possible and transpositions $[x, y]$ which convert a pre-lollipop to the double lollipop. There are two possible cases: x can be either a vertex of a cycle in the double-lollipop, or a vertex in the path connecting two cycles.

In the first case, the resulting graph is contained in the L neighborhood of the other cycle. In the second case, we might get two subgraphs each contained in the L -neighborhood of one of the cycles. See Figure 6 for a list of possibilities. Until $\bar{\sigma}$ there are at most $O(\sqrt{\log n})$ many cycles of length at most L that ever exist during $[0, S_0]$ with high probability. Thus, the total number of vertices contained in the L neighborhood of these cycles is also $O(\sqrt{\log n})$, where the big- O involves a constant that depends on d and L . Since pre-lollipops can only get born if the transposition involves vertices from these neighborhoods, there can be at most $O(\log n)$ many such pre-lollipops that ever exists during $[0, \bar{\sigma})$ with high probability. The rest of the argument is taken care of as in Lemma 21 and the discussion in the previous paragraph. \square

Lemma 27. *For any $T_0, S_0 > 0$, $K \in \mathbb{N}$ and any $\epsilon > 0$, for all T large enough we have*

$$\mathbb{P}(G(t, s) \text{ is } (K, K) \text{ tangle free for all } (t, s) \in R_0) \geq 1 - \epsilon,$$

where $R_0 := [T - T_0, T] \times [0, S_0]$.

Proof. This proof is very similar to the arguments in Section 4.4. If some $G(t, s)$ is not (K, K) tangle free then there exists two cycles \mathcal{C}_1 and \mathcal{C}_2 of size at most K which are at distance at most K . Moving forward in dimension, these cycles grow

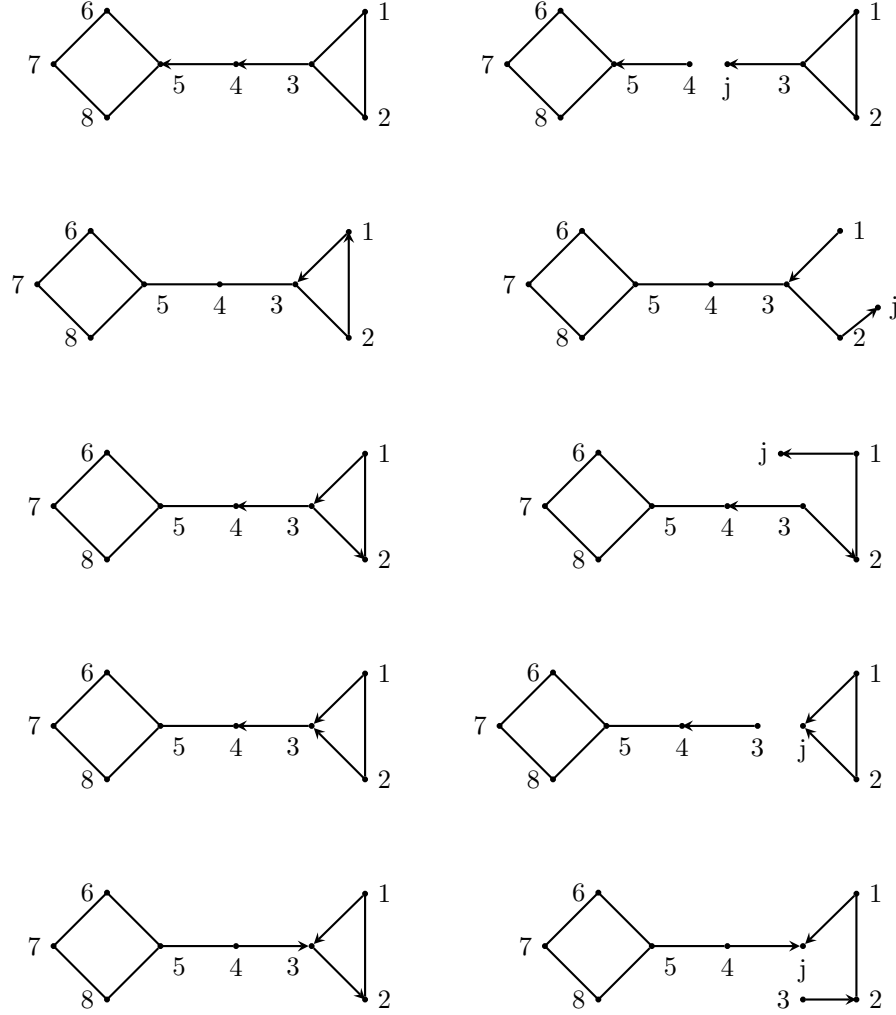


FIGURE 6. Distinct classes of pre-lollipops for the double-lollipop in Figure 5. Arrows represent the signs of the letters on the edges, and are only shown on the edges incident on the vertex that is relevant for the transposition. The vertex j refers to the other label in the transposition and is assumed to be not a vertex in the walk itself. In each case the resulting collection of trails is contained in the neighborhoods of one or more cycles. In all other cases either the graph does not change, except for vertex labels, or the pre-lollipop is similar to one above by symmetry.

into an overlap of cycles \tilde{C}_1, \tilde{C}_2 . By (45) there exists L such that with probability at least $1 - \varepsilon/8$ neither \tilde{C}_1 or \tilde{C}_2 contains a cycle of size bigger than L . Now Lemma 26 then implies that with probability at least $1 - \varepsilon/8$, \tilde{C}_1 and \tilde{C}_2 should contain exactly one cycle each and also should be disjoint. This is possible only if C_1 and

C_2 were disjoint and had a path \mathcal{P} joining them of length at most K in $G(t, s)$. Now growing in dimension \mathcal{P} will grow in to $\tilde{\mathcal{P}}$. Now some of the vertices on $\tilde{\mathcal{P}}$ might be repeated and also be the same as some of the vertices on either $\tilde{\mathcal{C}}_1$ or $\tilde{\mathcal{C}}_2$. However it is clear that one gets a double-lollipop at dimension T with two cycles $\tilde{\mathcal{C}}_2$ and $\tilde{\mathcal{C}}_2$ and a path \mathcal{P}_* joining them of length say ℓ . Since \mathcal{P} had length at most K it is clear that \mathcal{P}_* has at least $\ell - K$ shrinking vertices. Also both $\tilde{\mathcal{C}}_1$ and $\tilde{\mathcal{C}}_2$ have at least $|\tilde{\mathcal{C}}_1| - K$ and $|\tilde{\mathcal{C}}_2| - K$ shrinking vertices respectively. Again by Lemma 26 with probability at least $1 - \varepsilon/8$, we must have $\ell > L$.

Thus, we have shown that

$$\begin{aligned} P(G(t, s) \text{ is not } (K, K) \text{ tangle free}) &= \varepsilon/2 \\ &\leq P(\exists \text{ a lollipop of length } \ell > L \text{ with at least } \ell - 3K \text{ shrinking vertices}). \end{aligned}$$

Define the size of double-lollipop to be the length of the word associated to it.

Clearly now it suffices to show that there exists an L , such that the probability that any of the possible double-lollipops of size $l > L$ shrinks to a double-lollipop of size at most K by dimension $T - T_0$ is bounded above by $\varepsilon/4$.

Since the argument is very similar to that in Section 4.4, we simply point out the major differences. Consider (43). We will modify it so that $E_T(L)$ and $F(w, I)$ will refer to double-lollipops and not cycles. In (44), we modify G_w to refer to the number of double-lollipops that ever get created at dimension T during time $[0, S_0]$. Now similar arguments as in Lemma 22 gives us that

$$E(G_w) \leq c_0 S_0 |w|,$$

where c_0 is some absolute constant. Equation (17) adapted to this setting now shows that the number of possible w 's of length l which can shrink to a double-lollipop of size k is at most $l^k (2d)^k$.

Thus by application of union bound similar to (45) we are done. \square

Proof of Proposition 23. Follows from Lemmas 25 and 27. \square

Proof of Theorem 4. This is almost identical to the proof of Theorem 5 in [JP14]. We provide the sketch below. Since by Proposition 23 with probability approaching one, all CNBW's are repeated cycles, their weak limits can be computed from one another. Formally for any positive integer K , as $T \rightarrow \infty$,

$$(49) \quad \{\text{CNBW}_k(T + \cdot, \cdot) : 1 \leq k \leq K\} \xrightarrow{\text{weakly}} \left\{ \sum_{j|k} 2j N_j(\cdot, \cdot) : 1 \leq k \leq K \right\}.$$

The above follows from Theorem 1, Proposition 23, and easy applications of Continuous Mapping Theorem and Slutsky's theorem. Using (46) and (49) the polynomials f_k are obtained from Γ_k by expressing $N_k(\cdot, \cdot)$ as linear combinations of $\sum_{j|k} 2j N_j(\cdot, \cdot)$. This is done using the Möbius inversion formula. The inversion formula and the polynomial basis referred to is explicitly evaluated in [JP14, eqn. (16)]. \square

Proof of Theorem 5. By Proposition 23

$$2\text{tr} T_k(G(\infty + t, s)) = (2d - 1)^{-k/2} \sum_{j|k} 2j N_j(t, s).$$

Taking expectations on both sides, we get

$$2\mathbb{E}[\mathrm{tr}T_k(G(\infty + t, s))] = (2d - 1)^{-k/2} \sum_{j|k} a(d, j).$$

As d tends to infinity, from Theorem 2, only the N_k term in the difference survives. Thus, the field (10) has the same weak limit as one half of the field (7), which is $U/2$. This completes the proof of the proposition. \square

REFERENCES

- [Bil99] Patrick Billingsley. *Convergence of probability measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999. A Wiley-Interscience Publication.
- [BN88] Timothy C. Brown and Gopalan M. Nair. A Simple Proof of the Multivariate Random Time Change Theorem for Point Processes. *Journal of Applied Probability*, 25(1):210–214, 1988.
- [Bor10] Alexei Borodin. CLT for spectra of submatrices of Wigner random matrices II. Stochastic evolution. Preprint. Available at arXiv:1011.3544, 2010.
- [Bor14] Alexei Borodin. CLT for spectra of submatrices of Wigner random matrices. *Moscow mathematical journal*, 2014(1):29–38, 2014.
- [BSZ11] Nathanaël Berestycki, Oded Schramm, and Ofer Zeitouni. Mixing times for random k -cycles and coalescence-fragmentation chains. *The Annals of Probability*, 39(5):1815–1843, 09 2011.
- [Dia88] Persi Diaconis. *Group representations in probability and statistics*, volume 11 of *Lecture Notes-Monograph Series*. Institute of mathematical statistics, Hayward, 1988.
- [DJPP12] Ioana Dumitriu, Tobias Johnson, Soumik Pal, and Elliot Paquette. Functional limit theorems for random regular graphs. *Probability Theory and Related Fields*, pages 1–55, 2012. Published online, 25 August 2012.
- [DMWZZ04] Persi Diaconis, Eddy Mayer-Wolf, Ofer Zeitouni, and Martin P. W. Zerner. The poisson-dirichlet law is the unique invariant distribution for uniform split-merge transformations. *The Annals of Probability*, 32(1B):915–938, 01 2004.
- [DP12] Ioana Dumitriu and Soumik Pal. Sparse regular random graphs: Spectral density and eigenvectors. *Ann. Probab.*, 40(5):2197–2235, 2012.
- [DZ08] Youri Davydov and Ričardas Zitikas. On weak convergence of random fields. *Annals of the Institute of Statistical Mathematics*, 60(2):345–365, 2008.
- [Fri91] Joel Friedman. On the second eigenvalue and random walks in random d -regular graphs. *Combinatorica*, 11(4):331–362, 1991.
- [Fri08] Joel Friedman. A proof of Alon’s second eigenvalue conjecture and related problems. *Mem. Amer. Math. Soc.*, 195(910):viii+100, 2008.
- [GS84] R. K. Gettoor and M. J. Sharpe. Naturality, standardness, and weak duality for Markov processes. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, 67:1–62, 1984.
- [Joh14] Tobias Johnson. *Eigenvalue fluctuations for random regular graphs*. PhD thesis, University of Washington, 2014.
- [JP14] Tobias Johnson and Soumik Pal. Cycles and eigenvalues of sequentially growing random regular graphs. *The Annals of Probability*, 42(4):1396–1437, 2014.
- [KOV04] Sergei Kerov, Grigori Olshanski, and Anatoly Vershik. Harmonic analysis on the infinite symmetric group. *Invent. Math.*, 158(3):551–642, 2004.
- [KS91] I. Karatzas and S.E. Shreve. *Brownian motion and stochastic calculus (Second edition)*, volume 113 of *Graduate texts in mathematics*. Springer-Verlag, 1991.
- [KT75] S. Karlin and H.E. Taylor. *A First Course in Stochastic Processes*. Elsevier Science, 1975.
- [LP10] Nati Linial and Doron Puder. Word maps and spectra of random graph lifts. *Random Structures Algorithms*, 37(1):100–135, 2010.
- [Mey71] P. A. Meyer. Démonstration simplifiée d’un théorème Knight. In *Séminaire de Probabilités V*, volume 191 of *Springer lecture notes in mathematics*, pages 191–195. Springer, 1971.

- [Neu71] Georg Neuhaus. On weak convergence of stochastic processes with multidimensional time parameter. *The Annals of Mathematical Statistics*, 42(4):1285–1295, 08 1971.
- [Pit06] Jim Pitman. *Combinatorial stochastic processes*, volume 1875 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 2006. Lectures from the 32nd Summer School on Probability Theory held in Saint-Flour, July 7–24, 2002, With a foreword by Jean Picard.
- [Res07] S.I. Resnick. *Extreme Values, Regular Variation, and Point Processes*. Springer Series in Operations Research and Financial Engineering. Springer, 2007.
- [TVW13] Linh V. Tran, Van H. Vu, and Ke Wang. Sparse random graphs: Eigenvalues and eigenvectors. *Random Structures Algorithms*, 42(1):110–134, 2013.
- [Wat64] S. Watanabe. On discontinuous additive functionals and Lévy measures of a markov process. *Jap. J. Math.*, 34:31–70, 1964.

DEPARTMENT OF MATHEMATICS, UNIVERSITY OF WASHINGTON, SEATTLE, WA 98195

E-mail address: `sganguly@math.washington.edu`, `soumikpal@gmail.com`